

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-139404

(43)Date of publication of application : 20.05.1994

(51)Int.Cl.

G06K 9/20

(21)Application number : 04-310861

(71)Applicant : FUJI XEROX CO LTD

(22)Date of filing : 27.10.1992

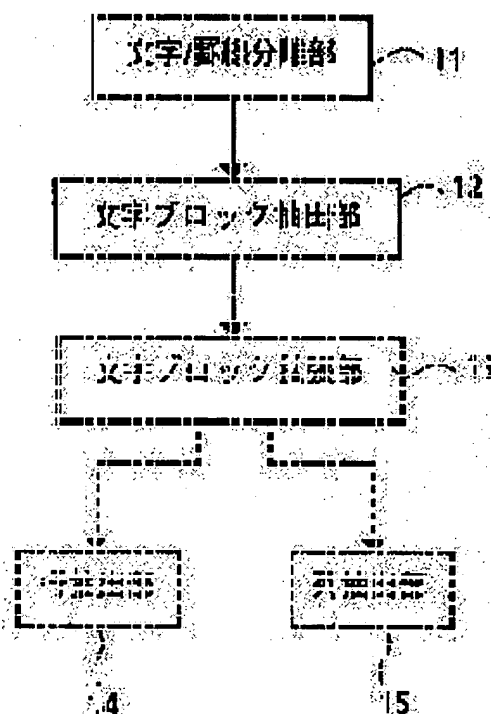
(72)Inventor : ITONORI KATSUHIKO

(54) TABLE RECOGNIZER

(57)Abstract:

PURPOSE: To precisely take out the structure of a table even in regard to such tables where the ruled lines are much omitted or the content is partly omitted.

CONSTITUTION: A table recognizer recognizes a table image including the characters and ruled lines and contains a character/ruled line separating part 11 which separates the characters from the ruled lines included in a table image, a character block extracting part 12 which extracts the character blocks from the character images separated by the part 11, a character block extending part 13 which extends evenly the edges of each character block based on the mutual position relations of those character blocks extracted by the part 12, and a row extracting part 14 which extracts the rows out of the character blocks based on the position relation among the character blocks extended by the part 13, and a column extracting part 15 which extracts the columns out of the character blocks based on the position relation among the character blocks extended by the part 13 respectively.



LEGAL STATUS

[Date of request for examination]

08.09.1995

[Date of sending the examiner's decision of

rejection]

[Kind of final disposal of application other than
the examiner's decision of rejection or
application converted registration]

[Date of final disposal for application]

[Patent number] 2789971

[Date of registration] 12.06.1998

[Number of appeal against examiner's
decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The alphabetic character / a ruled line separation means to separate the alphabetic character and ruled line in a table image in the table recognition equipment which recognizes the table image with which the alphabetic character and the ruled line are intermingled, An alphabetic block extract means to extract an alphabetic block from the alphabetic character image separated by said alphabetic character / ruled line separation means, Table recognition equipment characterized by having an alphabetic block escape means to extend each alphabetic block extracted by said alphabetic block extract means so that the edge of each alphabetic block may be arranged from mutual physical relationship.

[Claim 2] Table recognition equipment characterized by establishing a line sampling means by which the physical relationship between the alphabetic blocks extended by said alphabetic block escape means extracts a line in table recognition equipment according to claim 1, and a train extract means by which the physical relationship between the alphabetic blocks extended by said alphabetic block escape means extracts a train.

[Claim 3] The alphabetic character / a ruled line separation means to separate the alphabetic character and ruled line in a table image in the table recognition equipment which recognizes the table image with which the alphabetic character and the ruled line are intermingled, An alphabetic block extract means to extract an alphabetic block from the alphabetic character image separated by said alphabetic character / ruled line separation means, From the ruled line image separated by said alphabetic character / ruled line separation means, and each alphabetic block extracted by said alphabetic block extract means, an alphabetic block so that other alphabetic blocks may not be overlapped The 1st alphabetic block escape means extended to the nearest ruled line, and the 2nd alphabetic block escape means which extends each alphabetic block extended by said 1st alphabetic block escape means so that the edge of each alphabetic block may be arranged from mutual physical relationship, Table recognition equipment characterized by establishing a line sampling means by which the physical relationship between the alphabetic blocks extended by said 2nd alphabetic block escape means extracts a line, and a train extract means by which the physical relationship between the alphabetic blocks extended by said 2nd alphabetic block escape means extracts a train.

[Claim 4] The alphabetic character / a ruled line separation means to separate the alphabetic character and ruled line in a table image in the table recognition equipment which recognizes the table image with which the alphabetic character and the ruled line are intermingled, An alphabetic block extract means to extract an alphabetic block from the alphabetic character image separated by said alphabetic character / ruled line separation means, A rectangle frame extract means to extract the rectangle which consists of ruled line images separated by said alphabetic character / ruled line separation means by the ruled line, A configuration frame extract means to extract the rectangle frame which contains one or less alphabetic block according to the inclusion relation of the rectangle frame extracted with said rectangle frame extract means, and each alphabetic block extracted by the alphabetic block extract means as a frame which constitutes a table, The alphabetic block made into an object from the alphabetic block which is not contained in the configuration frame extracted by said configuration frame extract means, and the

ruled line image separated by said alphabetic character / ruled line separation means so that other alphabetic blocks may not be overlapped The 1st alphabetic block escape means extended to the nearest ruled line, and the 2nd alphabetic block escape means which extends each alphabetic block extended by said 1st alphabetic block escape means so that the edge of each alphabetic block may be arranged from mutual physical relationship, A line sampling means by which the physical relationship between the alphabetic block extended by said 2nd alphabetic block escape means and the configuration frame extracted by said configuration frame extract means extracts a line, Table recognition equipment characterized by having a train extract means by which the physical relationship between the alphabetic block extended by said 2nd alphabetic block escape means and the configuration frame extracted by said configuration frame extract means extracts a train.

[Claim 5] The alphabetic character / a ruled line separation means to separate the alphabetic character and ruled line in a table image in the table recognition equipment which recognizes the table image with which the alphabetic character and the ruled line are intermingled, An alphabetic block extract means to extract an alphabetic block from the alphabetic character image separated by said alphabetic character / ruled line separation means, From the ruled line image separated by said alphabetic character / ruled line separation means, and each alphabetic block extracted by the alphabetic block extract means, an alphabetic block so that other alphabetic blocks may not be overlapped The 1st alphabetic block escape means extended to the nearest ruled line, and the 2nd alphabetic block escape means which extends each alphabetic block extended by said 1st alphabetic block escape means so that the edge of each alphabetic block may be arranged from mutual physical relationship, An alphabetic block normalization means to normalize an alphabetic block according to the physical relationship of the alphabetic block extended by said 2nd alphabetic block escape means, An alphabetic block interpolation means to detect the rectangle which each alphabetic block which said alphabetic block normalization means normalized overlaps, and to consider that the rectangle is an imagination alphabetic block, Table recognition equipment characterized by having a line sampling means by which the physical relationship between the alphabetic blocks interpolated by said alphabetic block interpolation means extracts a line, and a train extract means by which the physical relationship between the alphabetic blocks interpolated by said alphabetic block interpolation means extracts a train.

[Translation done.]

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention recognizes the table image with which the alphabetic character and the graphic form are intermingled, and relates to the table recognition equipment which takes out the structure of a row and column.

[0002]

[Description of the Prior Art] As a method of the conventional table recognition, the ruled line which constitutes marginal distribution and the table of a table field is changed into a part for the vector line, and the rectangle frame surrounded by the ruled line is extracted. As a method which uses marginal distribution, there is a technique given in JP,1-129358,A etc. as a technique given [for example,] in JP,2-61775,A, and a method which uses a part for the vector line. The former method divides into two or more rectangle frames the ruled line of the outer frame which is in the outermost part of a table according to marginal distribution by the ruled line which touches ejection and this outer frame in ends. Furthermore, the rectangle frame surrounded by the ruled line is extracted by performing same processing recursively to within the limit [each / rectangle] which was divided. The latter method recognizes a table by investigating the physical relationship of each rectangle frame which pursued and took out a part for the vector line.

[0003] Although premised on these conventional methods not having an abbreviation in the ruled line which constitutes a table, even when the ruled line of both the sides of a table is omitted in JP,2-264386,A, the method which enabled it to take out a rectangle frame correctly is indicated. That is, this method is a method which generates a vertical ruled line virtually by both side of a table, when it distinguishes whether a ruled line is in both the sides of a table from the vertical ruled line taken out from the table image, and a horizontal ruled line and there is nothing.

[0004] The above-mentioned conventional method is a table with which all ruled lines have gathered or only the outermost ruled line is omitted. That is, it could apply only to the table of drawing 2 as shown in (a) and (b), and, in the case of a table like (c) - (e) of drawing 2, was not able to apply. The method which it is going to compensate with the ruled line currently omitted paying attention to the null field between character strings as a method applicable also to the table of drawing 2 as shown in (c) is indicated by JP,3-142691,A.

[0005]

[Problem(s) to be Solved by the Invention] However, as shown in a table (d) and (e), when nested structure existed in the line or the train, in order to find a continuous null field, the image needed to be investigated to the precision, and it had the fault to which the processing time becomes long. Moreover, also in recognition of the structure of the row and column of a table, the representation point (a core, center of gravity) was established into each element which constitutes a table, and the line between this representation point or the distance of the direction of a train was extracting the line and the train. That is, the line of all representation points or the distance of the direction of a train is investigated, and this distance extracts the following [a threshold] as one line or a train. However, when an alphabetic block

was used as the component of a table, and a gap of the location between alphabetic blocks was large, it was not necessarily what can extract a line and a train to accuracy.

[0006] Moreover, it was the main objects to start the part in which each alphabetic character exists so that OCR (character reader) could recognize the alphabetic character in a table image to accuracy, and the conventional method was not what saves the structure of table itself. Therefore, what was necessary was just to have been able to pinpoint the range as which the alphabetic character is filled in, even if the content of the table was omitted how. Conversely, when saying, since it did not need to input into OCR, the part to which the content of the table is abbreviated was able to be disregarded. However, the structure of a table is taken out, and when inputting into document preparation equipments, such as a word processor, again and using for them, the abbreviation of the content of the table may be unable to be disregarded. For example, in the case of the table described only by the horizontal ruled line as shown in the table of (e) of drawing 2, the abbreviation of the content caused trouble to the activity which extracts the line of a table, and the structure of a train, and had the inconvenience of it becoming impossible to take out the structure of a table to accuracy.

[0007] This invention aims at solving such a trouble. Namely, this invention aims at offering the table recognition equipment which can take out the structure of a table to accuracy also about a table with a part to which the ruled line of a table is abbreviated substantially, which exists a table and which is and is omitted by the content.

[0008]

[Means for Solving the Problem] In the table recognition equipment with which this invention (claim 1) recognizes the table image with which the alphabetic character and the ruled line are intermingled The alphabetic character in a table image, and the alphabetic character / ruled line separation means of separating a ruled line (11 of drawing 1), An alphabetic block extract means to extract an alphabetic block from the alphabetic character image separated by said alphabetic character / ruled line separation means (12 of drawing 1), It is characterized by having an alphabetic block escape means (13 of drawing 1) to extend each alphabetic block extracted by said alphabetic block extract means so that the edge of each alphabetic block may be arranged from mutual physical relationship.

[0009] This invention (claim 2) is characterized by to establish a line-sampling means (14 of drawing 1) to by_ which the physical relationship between the alphabetic blocks extended by said alphabetic block escape means extracts a line, and a train extract means (15 of drawing 1) to by_ which the physical relationship between the alphabetic blocks extended by said alphabetic block escape means extracts a train in table recognition equipment equipped with said alphabetic character / ruled line separation means, the alphabetic block extract means, and the alphabetic block extract means.

[0010] In said table recognition equipment with which this invention (claim 3) was equipped with said alphabetic character / ruled line separation means, the alphabetic block extract means, the alphabetic block escape means, the line sampling means, and the train extract means From the ruled line image separated by the alphabetic character / ruled line separation means (151 of drawing 15), and each alphabetic block extracted by the alphabetic block extract means (152 of drawing 15), an alphabetic block escape means (153 of drawing 15) an alphabetic block so that other alphabetic blocks may not be overlapped The 1st alphabetic block escape means extended to the nearest ruled line (1531 of drawing 15), It has the 2nd alphabetic block escape means (1532 of drawing 15) which extends each alphabetic block extended by the 1st alphabetic block escape means so that the edge of each alphabetic block may be arranged from mutual physical relationship.

[0011] In the table recognition equipment with which this invention (claim 4) recognizes the table image with which the alphabetic character and the ruled line are intermingled The alphabetic character in a table image, and the alphabetic character / ruled line separation means of separating a ruled line (211 of drawing 21), An alphabetic block extract means to extract an alphabetic block from the alphabetic character image separated by said alphabetic character / ruled line separation means (212 of drawing 21), A rectangle frame extract means to extract the rectangle which consists of ruled line images separated by said alphabetic character / ruled line separation means by the ruled line (213 of drawing 21), With the rectangle frame and alphabetic block extract means which were extracted with said

rectangle frame extract means A configuration frame extract means to extract the rectangle frame (namely, rectangle frame which does not contain the rectangle frame and alphabetic block only containing one alphabetic block) which contains one or less alphabetic block according to inclusion relation with each extracted alphabetic block as a frame which constitutes a table (214 of drawing 21), The alphabetic block made into an object from the alphabetic block which is not contained in the configuration frame extracted by said configuration frame extract means, and the ruled line image separated by said alphabetic character / ruled line separation means so that other alphabetic blocks may not be overlapped The 1st alphabetic block escape means extended to the nearest ruled line (2151 of drawing 21), The 2nd alphabetic block escape means which extends each alphabetic block extended by said 1st alphabetic block escape means so that the edge of each alphabetic block may be arranged from mutual physical relationship (2152 of drawing 21), A line sampling means by which the physical relationship between the alphabetic block extended by said 2nd alphabetic block escape means and the configuration frame extracted by said configuration frame extract means extracts a line (216 of drawing 21), It is characterized by having a train extract means (217 of drawing 21) by which the physical relationship between the alphabetic block extended by said 2nd alphabetic block escape means and the configuration frame extracted by said configuration frame extract means extracts a train.

[0012] In the table recognition equipment with which this invention (claim 5) recognizes the table image with which the alphabetic character and the ruled line are intermingled The alphabetic character in a table image, and the alphabetic character / ruled line separation means of separating a ruled line (281 of drawing 28), An alphabetic block extract means to extract an alphabetic block from the alphabetic character image separated by said alphabetic character / ruled line separation means (282 of drawing 28), From the ruled line image separated by said alphabetic character / ruled line separation means, and each alphabetic block extracted by the alphabetic block extract means, an alphabetic block so that other alphabetic blocks may not be overlapped The 1st alphabetic block escape means extended to the nearest ruled line (283 of drawing 28), The 2nd alphabetic block escape means which extends each alphabetic block extended by said 1st alphabetic block escape means so that the edge of each alphabetic block may be arranged from mutual physical relationship (284 of drawing 28), An alphabetic block normalization means to normalize an alphabetic block according to the physical relationship of the alphabetic block extended by said 2nd alphabetic block escape means (285 of drawing 28), An alphabetic block interpolation means to detect the rectangle which each alphabetic block which said alphabetic block normalization means normalized overlaps, and to consider that the rectangle is an imagination alphabetic block (286 of drawing 28), A line sampling means by which the physical relationship between the alphabetic blocks interpolated by said alphabetic block interpolation means extracts a line (287 of drawing 28), It is characterized by having a train extract means (288 of drawing 28) by which the physical relationship between the alphabetic blocks interpolated by said alphabetic block interpolation means extracts a train.

[0013]

[Function] In an operation of this invention (claim 1), an alphabetic character / ruled line separation means separates the alphabetic character and ruled line which exist in a table image, and it extends with an alphabetic block escape means so that the edge of all the alphabetic blocks for which asked for the alphabetic block of 1 settlement with the alphabetic block extract means from the distance between the black pixel lumps in the alphabetic character image for which it asked etc., and it asked with this alphabetic block extract means may be arranged. Thus, since a table is recognized by burying the gap which this invention extends the alphabetic block which constitutes a table, and is between alphabetic blocks, the structure can be recognized also about a table with which the ruled line of a table is omitted substantially, and even if the location gap between alphabetic blocks is large, moreover, exact recognition can be performed.

[0014] In this invention (claim 2), a line sampling means and a train extract means investigate the list of the line writing direction of the alphabetic block which the above-mentioned extended, and the direction of a train, and the structure of a line and a train is taken out. Since the block which extended the alphabetic block which constitutes a table from this invention extracts the list of the line writing

direction of a table, and the direction of a train, the structure of a table can be extracted to accuracy. [0015] In this invention (claim 3), it extends with the 2nd alphabetic block escape means so that the edge of each alphabetic block which extended each alphabetic block with the 1st alphabetic block escape means to the nearest ruled line so that other alphabetic blocks might be straddled for no alphabetic blocks for which it asked with the alphabetic block extract means using the ruled line image for which it asked with the alphabetic character / ruled line separation means, and was extended by this 1st alphabetic block escape means 203 may be arranged. According to this, since the information on a ruled line is used for the escape of an alphabetic block, the structure of a table can be extracted more to accuracy. For example, even if it is a table (drawing 20 (a)) with the alphabetic block over two or more division trains like [for identification division / the divided train], the structure of a table can be acquired to accuracy (drawing 20 (c)).

[0016] In this invention (claim 4), the rectangle frame extract means extracts all the rectangles formed by the ruled line from the ruled line image separated with the alphabetic character / ruled line separation means. With a configuration frame extract means, the inclusion relation of the alphabetic block extracted by the alphabetic block extract means and the rectangle extracted with the rectangle frame extract means is investigated, and let the rectangle (namely, rectangle which does not contain one alphabetic block **** rectangle or an alphabetic block) containing one or less alphabetic block be the configuration frame of a table. A line sampling means and a train extract means investigate the list of the line writing direction of the configuration frame which extended the alphabetic block with the 1st alphabetic block escape means and the 2nd alphabetic block escape means, and asked for the alphabetic block which is not contained in the configuration frame of a table with this alphabetic block and the above-mentioned configuration frame extract means, and the direction of a train, and the structure of a line and a train is taken out. Since according to this invention the rectangle formed by the ruled line of a table is used in order to grasp the structure of a table, even if it is the table of a complicated configuration, the structure can be extracted to accuracy.

[0017] In this invention (claim 5), an alphabetic character / ruled line separation means 1 separates the alphabetic character and ruled line which exist in a table image. It asks for the alphabetic block of 1 settlement with an alphabetic block extract means from the distance between the black pixel lumps in the alphabetic character image for which it asked. Each alphabetic block is extended with the 1st alphabetic block escape means to the nearest ruled line so that other alphabetic blocks may be straddled for no alphabetic blocks for which it asked with the alphabetic block extract means using the ruled line image for which it asked with the alphabetic character / ruled line separation means. It extends with the 2nd alphabetic block escape means so that the edge of all alphabetic blocks may furthermore be arranged. Next, this extended alphabetic block is normalized in the location of an alphabetic block, and magnitude is normalized with an alphabetic block normalization means from the information on that rectangle field. With an alphabetic block interpolation means, duplication between this alphabetic block that normalized is investigated, an imagination alphabetic block is prepared in a duplicate part, a line sampling means and a train extract means investigate the list of the line writing direction of an alphabetic block, and the direction of a train, and the structure of a line and a train is taken out. According to this invention, even if it is the table which has an abbreviation in the content by extending an alphabetic block so that the edge may be arranged, and normalizing a location and magnitude, the structure of a table can be extracted to accuracy.

[0018]

[Example]

The 1st example drawing 1 is drawing showing the configuration of the 1st example of this invention. This equipment is equipped with an alphabetic character / ruled line separation section 11, the alphabetic block extract section 12, the alphabetic block extension 13, the line sampling section 14, and the train extract section 15. An alphabetic character / ruled line separation section 11 performs processing which separates the alphabetic character currently written into the table image, and a ruled line. This processing is realizable by investigating the area of the lump of the pixel in the image which forms an alphabetic character, and the lump of the pixel which forms a ruled line (graphic form), a profile,

complexity, etc. For example, the well-known technique indicated by "1 examination of the production system installation in an alphabetic character and graphic form separation processing" (PRU 83-62 P.67-74) of Iwaki and others can be used. Moreover, it is more suitable if a technique given [by these people] in Japanese Patent Application No. No. 290299 [three to] is used. Paying attention to an alphabetic character image, processing is performed among the images with which the processing after this was separated.

[0019] In the alphabetic block extract section 2, about the alphabetic character image obtained by above-mentioned alphabetic character / ruled line separation section 11, it asks for a rectangle field including one pixel lump, this is presumed to be one alphabetic character, one or more alphabetic characters which are approaching with the distance between alphabetic characters are packed, and it unifies as an alphabetic block. Since Japanese consists of two or more pixel lumps in many cases as shown in drawing 3, it is wrong as original semantics to make one pixel lump correspond to one alphabetic character simply, but like drawing 3 (a), since these pixel lumps are approaching dramatically, they do not become a problem by this example. Therefore, special processing is not performed here. However, when there is a part which two pixel lumps overlap in that rectangle field like drawing 3 (b), suppose that these two pixel lumps are unified and a rectangle field is newly set up. In addition, when it is necessary to ask accuracy at a time for one alphabetic character more, a technique given in JP,3-267278,A can be used.

[0020] The alphabetic block extract section is explained in detail using the flow of drawing 4 and drawing 5. However, the rectangle field of the pixel lump which expresses all the alphabetic characters in an alphabetic character image at this event shall be called for, and this rectangle field is expressed as an alphabetic character rectangle. First, in drawing 4, it asks for total of the magnitude of each alphabetic character rectangle (steps 401-404), and considers as the thresholds T_w and T_h when unifying one half of the averages of the magnitude of total as an alphabetic block. Although the one half of the average magnitude of an alphabetic character rectangle is decided as a threshold here, how to decide this threshold may take approaches, such as deciding as several% of the average of the distance not only between this but alphabetic character rectangles.

[0021] Next, in drawing 5, every one alphabetic character rectangle is investigated [ejection and] for whether it is already registered as a part of alphabetic block (step 407). If it is not a part of alphabetic block, an alphabetic block will newly be made and this alphabetic block will be registered as an element of that head (step 408). Next, it investigates whether it is smaller than the thresholds T_w and T_h which found distance with this alphabetic block about the remaining alphabetic character rectangles (step 411), and the distance of level and a perpendicular direction calculated by drawing 4 (step 412). In being small, it registers as a part of alphabetic block (step 413). The above processing is repeated until the alphabetic character rectangle which is not registered into an alphabetic block is lost.

[0022] The above processing can extract an alphabetic block, as shown in drawing 6 (a), (b), drawing 7 (a), and (b). Although it investigates only about an alphabetic character image and the alphabetic block is extracted here, however the pixel lump showing an alphabetic character may approach and exist using the ruled line image for which it can ask in an alphabetic character / ruled line separation section 1, it will become possible to ask for a more exact alphabetic block by forbidding what it straddles and unifies a ruled line.

[0023] Next, the alphabetic block extension 13 investigates distribution of the alphabetic block in the lengthwise direction of each alphabetic block, and a longitudinal direction, and it extends each alphabetic block so that the edge may be arranged. Processing of the alphabetic block extension 13 is explained using the flow of drawing 8. First, this processing consists of an escape to a longitudinal direction, and an escape to a lengthwise direction. Alphabetic block distribution of the lengthwise direction showing with which coordinate each alphabetic block has lapped how many is created (step 801). The alphabetic block distribution 91 of the lengthwise direction to the table of drawing 6 (a) is shown in drawing 9. In addition, the lateral alphabetic block distribution 92 is also doubled and shown in drawing 9. Next, Variable i is reset to 0 (step 802). The value of the alphabetic block distribution $HistW$ of a lengthwise direction changes to i , 0 [i.e.,], or the part which changes to other larger values

than 0-0 is found, and the width of face of each alphabetic block is extended so that other alphabetic blocks may not be straddled (step 803). However, it extends by setting up the following two rules at this time.

rule 1: others -- the time of changing from a value to i -- extended rule 2 to the right: When changing to other larger values than i to i , processing of ***** to the left is continued while being $i < N$. The situation of the alphabetic block in the event of the escape of the width of face of an alphabetic block ending the result of step 803 at the time of $i = 0$ to drawing 10 (a) is shown in drawing 10 (b).

[0024] Next, lateral alphabetic block distribution is created and height is extended by same processing (steps 804-806). in addition, the rule of an escape at step 806 -- rule 1: others -- the time of changing from a value to i -- down extended rule 2: When changing to other larger values than i to i , the alphabetic block to which the edge is equal in each line and a train as shown in drawing 11 with the processing beyond the escape to above can be obtained.

[0025] In the line sampling section 14 and the train extract section 15, the relation of a row and column is extracted by investigating the physical relationship of the alphabetic block extended by the alphabetic block extension 13. Even if it sees alphabetic block distribution of drawing 11, each line and the alphabetic block which constitutes each train exist within a certain fixed limits so that clearly. Therefore, it is possible to extract each line and each train to accuracy by investigating this inclusion relation. The processing flow of the line sampling section 14 is shown in drawing 12, and the processing flow of the train extract section 15 is shown in drawing 13.

[0026] In the line sampling section 14, each alphabetic block is first rearranged into ascending order by the Y coordinate of the top chord of an alphabetic block (step 1201). Next, the alphabetic block which exists within the limits of the Y coordinate of the Y coordinate of the range of the lengthwise direction in which an alphabetic block B_i exists the i -th alphabetic block B_i in the alphabetic block which is not registered into ejection (it resets to $i = 0$ at first) and a line, i.e., the top chord of an alphabetic block B_i , - the bottom side is registered as one line (step 1202). This processing is repeated until a non-registered alphabetic block is lost.

[0027] Moreover, the train extract section 15 rearranges each alphabetic block into ascending order by the X coordinate of the left part of an alphabetic block similarly (step 1301). Next, the alphabetic block which exists within the limits of the X coordinate of the X coordinate of the range of the longitudinal direction in which an alphabetic block B_i exists the i -th alphabetic block B_i in the alphabetic block which is not registered into ejection (it resets to $i = 0$ at first) and a train, i.e., the left part of an alphabetic block B_i , - the right-hand side is registered as one train (step 1302). This processing is repeated until a non-registered alphabetic block is lost.

[0028] Furthermore, as shown in the table of drawing 6 (a), when a line or a train has division, a line and a train can be extracted by performing processing of repeat line sampling and a train extract to each line and a train until two or more lines or trains are no longer extracted. The result which carried out line sampling to the table of drawing 6 (a) is shown in drawing 14 (a), and the result of a train extract is shown in drawing 14 (b). As mentioned above, the structure of the row and column in a table can be extracted to accuracy by extending an alphabetic block so that the edge may be arranged.

[0029] In addition to the 1st example, the 2nd example [2nd] of an example simplifies extended processing of an alphabetic block using the ruled line in a table image, and extracts the structure of a line and a train to accuracy more. Drawing 15 is drawing showing the basic configuration of the 2nd example. The table recognition equipment of this 2nd example comes to have an alphabetic character / ruled line separation section 151, the alphabetic block extract section 152, the alphabetic block extension 153, the line sampling section 154, and the train extract section 155, and that alphabetic block extension 153 consists of the 1st alphabetic block extension 1531 and the 2nd alphabetic block extension 1532. Since it is as the same as the example 1 has described, an alphabetic character / ruled line separation section 151, and the alphabetic block extract section 152 are not explained here.

[0030] The 1st alphabetic block extension 1531 extends an alphabetic character block size using a ruled line by considering as an input the alphabetic block extracted by the ruled line image separated by an alphabetic character / ruled line separation section 151 and the alphabetic block extract section 152. The

processing flow of the 1st alphabetic block extension 1531 is shown in drawing 16 and drawing 17. Order is explained for that processing later on using this flow. A ruled line image is vectorized (step 1601). The technique indicated by JP,1-142880,A, JP,2-105265,A, etc. can be used for the technique of this vectorization. Next, the distance of each alphabetic block and vector data is found, and it asks for the nearest vector data on the four directions of an alphabetic block, respectively (steps 1602, 1605, 1608, and 1611). Only when other alphabetic blocks cannot be found between the vector data and the alphabetic blocks for which it asked here, an alphabetic block is extended to it to vector data (step 1603-1604, 1606-1607, 1609-1610, 1612-1613). If the escape at this time is possible, it will perform an escape which makes in agreement the endpoint of vector data, and the angle of an alphabetic block. This processing is performed to all alphabetic blocks.

[0031] The processing result of the 1st [to the table of drawing 6 (a)] alphabetic block extension 1531 is shown in drawing 18 (a). Moreover, the processing result of the 1st alphabetic block extension 1531 is shown in drawing 19 together with all of all the ruled lines about the table which is. When all ruled lines have gathered so that clearly from drawing 19 $R > 9$, the alphabetic block called for by the 1st alphabetic block extension 1531 is in agreement with the rectangle surrounded by the ruled line. Since the edge of an alphabetic block has already gathered when such, processing of the 2nd alphabetic block extension 1532 is omissible. Next, the same technique as the alphabetic block extension 13 of the 1st example can be used for the 2nd alphabetic block extension 1532. The processing result of the 2nd [to the table of drawing 6 (a)] alphabetic block extension 1532 is shown in drawing 1818 (b).

[0032] Since it is the same processing as having explained the line sampling section 154 and the train extract section 155 in the 1st example, explanation is omitted here. A part of ruled line as shown in drawing 20 (a) is missing with the above processing, and it becomes possible to recognize the table with which the train is moreover divided. Although a mistaken result like drawing 20 (b) is brought in the alphabetic block extension 13 of the 1st example, the result which was in agreement with the structure of a table like drawing 20 (c) with the effectiveness of a block escape using the ruled line information on the 1st alphabetic block extension 1531 can be searched for.

[0033] The 3rd example [3rd] of an example performs extract processing of a line and a train to accuracy also using the rectangle made by the ruled line in a table image. Drawing 21 is drawing showing the basic configuration of the 3rd example. The table recognition equipment of this example is equipped with an alphabetic character / ruled line separation section 211, the alphabetic block extract section 212, the rectangle frame extract section 213, the configuration frame extract section 214, the alphabetic block extension 215, the line sampling section 216, and the train extract section 217, and the alphabetic block extension 215 consists of the 1st alphabetic block extension 2151 and the 2nd alphabetic block extension 2152 further. The inputted table image is divided into an alphabetic character image and a ruled line image by an alphabetic character / ruled line separation section 211. The alphabetic block extract section 212 extracts the alphabetic block in a table by considering the separated alphabetic character image as an input. An alphabetic character / ruled line separation section 211, and the alphabetic block extract section 212 are the same as what was explained in the 1st example, and omit explanation here.

[0034] The rectangle frame extract section 213 considers a ruled line image as an input, and performs processing which extracts the rectangle formed by the ruled line. After this processing vectorizes a ruled line image, it investigates the relation between vector data and extracts the rectangle. Since the technique of this vectorization is a well-known technique, it does not state especially here (for example, refer to JP,1-142880,A and JP,2-105265,A). Since vertical vector data connected with right and left of one level vector data and level vector data has connected with the bottom of it further, the frame enclosed by the ruled line investigates each level vector data, and extracts it by the approach of registering as a frame the vector data which fulfills conditions.

[0035] It explains using the flow chart which shows this processing to drawing 22 and drawing 23. First, counting of the number of all the vector data that constitutes a table is carried out (step 2201). Processing of step 2202 to the following step 2212 is applied to all vector data. Next, the level vector data V_i used as a rectangle frame top ruled line is looked for (step 2203). This can find level vector data

from it being below a threshold with the include angle of vector data and a horizontal line to make. Since the level vector data V_i found here may serve as a k -th rectangle frame top ruled line, this vector data V_i is registered into the column of the k -th rectangle frame top ruled line of the rectangle frame configuration table 241 (step 2204). Next, the vector data which constitutes the side on the right-hand side of the rectangle frame W_k is looked for (step 2205). That is, processing which finds vertical vector data which has the endpoint of the way which touches the endpoint at the right end of vector data V_i , and is not in contact with vector data V_i below vector data V_i is performed. It can ask for vertical vector data easily from it being below a threshold with the include angle with a perpendicular to make. Since the vector data found at this step may constitute the right ruled line of the rectangle frame W_k , it registers with the column of the right ruled line of the k -th rectangle frame of the rectangle frame configuration table 241 (step 2206). At this time, it investigates whether the vector data extended on left-hand side has connected to this vector data twist under the vector data found as a right ruled line. Since the vertical vector data which touches the bottom of it may also constitute the right ruled line of the rectangle frame W_k when such vector data does not exist, it registers with the column of the right ruled line of the k -th rectangle frame of the rectangle frame configuration table 241.

[0036] Similarly, the left ruled line of the rectangle frame W_k is looked for (step 2207), and it registers with the column of the left ruled line of the k -th rectangle frame of the rectangle frame configuration table 241 (step 2208). Furthermore, level vector data which connects the right ruled line for which it asked now, and a left ruled line is found (step 2209), and it registers with the column of the bottom ruled line of the k -th rectangle frame of the rectangle frame configuration table 241 (step 2210). When a ruled line does not find at least one of the above processings, all registration of the k -th rectangle frame of the rectangle frame configuration table 241 is canceled, and it resets so that the rectangle frame which consists of other vector data can be registered. The rectangle frame configuration table 241 when applying the above processing to the table of drawing 24 (a) becomes like drawing 24 (b).

[0037] In the configuration frame extract section 214, the inclusion relation of the rectangle frame for which it asked in the alphabetic block and the rectangle frame extract section 213 for which it asked in the alphabetic block extract section 212 is investigated, an alphabetic block is included only one or the rectangle frame which does not contain an alphabetic block at all is extracted as a component of a table. The processing flow of the configuration frame extract section 214 is shown in drawing 25. The outline of processing investigates the inclusion relation of each rectangle frame and an alphabetic block (step 2503), and carries out counting of the number of the alphabetic blocks contained in a rectangle frame (step 2504). Consequently, since, as for the rectangle frame containing two or more alphabetic blocks, the structure of a row and column is governed by the alphabetic block of that interior, such a rectangle frame is rejected. Moreover, the rectangle frame containing one or less alphabetic block is registered as a configuration frame (steps 2507 and 2508). The result of the configuration frame extract section 214 about a table as shown in drawing 2626 (a) becomes like drawing 26 (b).

[0038] In the alphabetic block extension 215, by considering the ruled line image called for in an alphabetic character / ruled line separation section 211, the configuration frame called for by the configuration frame extract section 214, and the alphabetic block out of it as an input, it extends by the 1st alphabetic block extension 2151 and the 2nd alphabetic block extension 2152 so that the edge of an alphabetic block may gather. Since the 1st example and the 2nd example explain this 1st and 2nd alphabetic block extension, it omits explanation here. The line sampling section 216 and the train extract section 217 are the same as the processing which considering as an input only differ and explained the alphabetic block and the configuration frame in the 1st example. The configuration of an example 3 shows the result of having extracted the structure of a row and column to drawing 27 (a) and (b) from a table which is shown in drawing 26 (a). As explained above, even if it is various complicated tables, according to this example, it is possible to extract the structure of the line and a train to accuracy.

[0039] The 4th example drawing 28 is drawing showing the configuration of the 4th example of this invention. This is equipped with an alphabetic character / ruled line separation section 281, the alphabetic block extract section 282, the 1st alphabetic block extension 283, the 2nd alphabetic block extension 284, the alphabetic block normalization section 285, the alphabetic block interpolation section

286, the line sampling section 287, and the train extract section 288. Moreover, the 1st alphabetic block extension 283 consists of the 1st alphabetic block width-of-face extension 2831 and the 1st alphabetic block high extension 2832, the 2nd alphabetic block extension 284 consists of the 2nd alphabetic block width-of-face extension 2841, the 2nd alphabetic block high extension 2842, and the alphabetic block integrated section 2843, and the alphabetic block normalization section 285 consists of the alphabetic block width-of-face normalization section 2851 and the alphabetic block high normalization section 2852.

[0040] An alphabetic character / ruled line separation section 281 performs processing which separates the alphabetic character currently written into the table image, and a ruled line. About the alphabetic character image obtained by above-mentioned alphabetic character / ruled line separation section 1, the alphabetic block extract section 282 asks for a rectangle field including one black pixel lump, presumes this to be one alphabetic character, packs one or more alphabetic characters which are approaching with the distance between alphabetic characters, and unifies them as an alphabetic block. These are the same as what was explained in the 1st example, and omit explanation here.

[0041] Next, the 1st alphabetic block extension 283 extends an alphabetic character block size using a ruled line by considering as an input the alphabetic block extracted by the ruled line image separated by an alphabetic character / ruled line separation section 281 and the alphabetic block extract section 282. This 1st alphabetic block extension 283 consists of the 1st alphabetic block high extension 2832 which extends an alphabetic block in the vertical direction, and the 1st alphabetic block width-of-face extension 2831 which extends an alphabetic block to a longitudinal direction.

[0042] The algorithm which extends an alphabetic block in the vertical direction by the 1st alphabetic block high extension 2832 is explained using drawing 2929. First, a ruled line image is vectorized (step 291). The existing technique indicated by JP,1-142880,A, JP,2-105265,A, etc. can be used for the technique of this vectorization. Next, the distance of each alphabetic block and vector data is found, and it asks for the nearest vector data by the upper and lower sides of an alphabetic block, respectively. Only when other alphabetic blocks cannot be found between the vector data and the alphabetic blocks for which it asked here, an alphabetic block is extended to it to vector data (steps 292-297). If the escape at this time is possible, it will perform an escape which makes in agreement the endpoint of vector data, and the angle of an alphabetic block. This processing is performed to all alphabetic blocks. Drawing 30 shows similarly the algorithm which extends an alphabetic block to the longitudinal direction by the 1st alphabetic block width-of-face extension 2831. This algorithm considers that above [which were explained by above-mentioned drawing 29 / the ARIGO rhythm and above] are the left, and is [down] wholly the same as that of a thing the bottom in the right. The processing result of the 1st [to the table of drawing 6 (a)] alphabetic block extension 283 is shown in drawing 18 (a). Moreover, the processing result of the 1st alphabetic block extension 283 is shown in drawing 19 (b) about the table which is together with all of all the ruled lines as shown in drawing 1919 (a). When all ruled lines have gathered so that clearly from this drawing, the alphabetic block called for by the 1st alphabetic block extension 283 is in agreement with the rectangle surrounded by the ruled line.

[0043] Next, the 2nd alphabetic block extension 284 investigates distribution of the alphabetic block in a longitudinal direction, and the distribution in the lengthwise direction of each alphabetic block, it extends each alphabetic block so that the edge may be arranged, and it is constituted by the alphabetic block integrated section 1843 which unifies each result with the 2nd alphabetic block width-of-face extension 2841 and the 2nd alphabetic block high extension 2842.

[0044] Order is explained for each processing later on. First, processing of the 2nd alphabetic block width-of-face extension 2841 and the 2nd alphabetic block high extension 2842 is explained using the flow of drawing 31 and drawing 32. Each of these two processings are performed to juxtaposition. Alphabetic block distribution of the lengthwise direction which means with which coordinate each alphabetic block has lapped how many in step 3101 in the 2nd alphabetic block width-of-face extension 2841 is created. The alphabetic block distribution 331 of the lengthwise direction to the table of drawing 6 (a) is shown in drawing 33. In addition, the lateral alphabetic block distribution 332 is also doubled and shown in drawing 33 R> 3. Here, it is processing to the result of the 1st alphabetic block extension

283. Next, Variable i is reset to 0 at step 3102. At step 3103, the value of the alphabetic block distribution HistW of a lengthwise direction changes to i , 0 [i.e.,], or the part which changes to other larger values than 0-0 is found, and the width of face of each alphabetic block is extended so that other alphabetic blocks may not be straddled. However, it extends by setting up the following two rules at this time.

rule 1: others -- the time of changing from a value to i -- extended rule 2 to the right: When changing to other larger values than i to i , processing of ***** to the left is continued while being $i < N$. The situation of the alphabetic block in the event of the escape of the width of face of an alphabetic block being completed is shown in drawing 34 (a).

[0045] In addition, about the 2nd alphabetic block high extension 2842, in steps 3201-3203 of drawing 32, lateral alphabetic block distribution is created and height is extended by the same processing as the escape of above-mentioned width of face. in addition, the rule of an escape at step 3203 -- rule 1: others -- the time of changing from a value to i -- down extended rule 2: When changing to other larger values than i to i , it is the escape to above. The situation of an alphabetic block when the escape of height is completed to drawing 34 (b) is shown.

[0046] In the alphabetic block integrated section 2843, the result of the 2nd alphabetic block width-of-face extension 2841 and the 2nd alphabetic block high extension 2842 is considered as an input, and the alphabetic block in the condition that doubled each width of face and height and the edge of each alphabetic block gathered eventually is obtained. The width of face of the alphabetic block which this processing investigates each result and corresponds uses the processing result of the 2nd alphabetic block width-of-face extension 2841, height uses the processing result of the 2nd alphabetic block high extension 2842, and each alphabetic character block size is changed. The result of the alphabetic block integrated processing about the table of drawing 6 (a) is shown in drawing 35. Moreover, the processing result of the 2nd alphabetic block extension 284 about the table which has an abbreviation by the content in drawing 36 is shown. An alphabetic block overlaps in the part (slash) which has an abbreviation in the content so that drawing 36 (b) may see.

[0047] The location of all alphabetic blocks, width of face, and height are normalized from the positional information of the borderline of the four directions which constitute it for the alphabetic block extended by the 2nd above-mentioned alphabetic block extension 284 from the alphabetic block normalization section 285. This processing consists of the alphabetic block width-of-face normalization section 2851 and the alphabetic block high normalization section 2852, as shown in drawing 28. Here, the alphabetic block width-of-face normalization section 2851 is explained using drawing 37. First, all frames are investigated and it memorizes in Array BXR in quest of a right end X coordinate (step 3701). At this time, the overlapping X coordinate is not memorized in Array BXR. Moreover, it memorizes in Array BXL similarly about a left-hand side X coordinate (step 3702). At this time, the value of -1 is put into the element BXR of the beginning of Array BXR [1], and a actual coordinate value is memorized from the 2nd element BXR of Array BXR [2] in distinction from other coordinate values. Next, these two arrays BXL and BXR are sorted in ascending order (step 3703). Next, the X coordinate at the left end of all alphabetic blocks is investigated again, and the element number (the subscript of an array corresponds) of the array BXL to which the value corresponds is registered into the column of X of the alphabetic block normalization table 381 (step 3704). For example, in drawing 36, since the left coordinate of frame No.W1 corresponds to the 2nd element of Array BXL, 2 is registered into the column of X of frame No.W1 of the alphabetic block normalization table 381. Next, the X coordinate at the right end of all alphabetic blocks is investigated, and the value which lengthened the value of the column of X of the frame with which the alphabetic block normalization table 381 registered previously corresponds from the element number of the array BXR to which the value corresponds is registered into the column of W (width of face) of the alphabetic block normalization table 381 (step 3705). For example, in drawing 36, since the right coordinate of frame No.W1 corresponds to the 3rd element of Array BXR, 2 is registered into the column of W of frame No.W1 of the alphabetic block normalization table 381 (width of face). The alphabetic block high normalization section 2852 is realizable by performing similarly that the alphabetic block width-of-face normalization section 2851 is following the

X coordinate about a Y coordinate. The alphabetic block normalization table 381 about drawing 36 is shown in drawing 3838.

[0048] In the alphabetic block interpolation section 286, duplication of an alphabetic block is investigated based on the alphabetic block normalization table 381 for which it asked in the alphabetic block normalization section 285. The procedure of this processing is explained using the flow of drawing 39. First, initial value 0 and 1 is memorized to Variables i and j, respectively (step 3901). The i-th and the j-th alphabetic block are compared, and it investigates whether there is any part which overlaps in the alphabetic block normalization table 381 (step 3902). If there is a part which overlaps at this time, the imagination alphabetic block corresponding to a duplication part will be registered into the alphabetic block normalization table 381 (step 3903). Next, the part which overlaps from the i-th and the j-th alphabetic block is deleted, and it re(step 3904) registers with the alphabetic block normalization table 381. The above processing is performed to all alphabetic blocks. Signs that the alphabetic block normalization table 381 of drawing 38 processed as mentioned above was matched with the actual table again at drawing 40 (a) are shown in drawing 40 (b). An imagination alphabetic block is generable to a part for the content block currently omitted with the above processing.

[0049] In the line sampling section 287 and the train extract section 288, the alphabetic block normalization section 285 normalizes and the relation of a row and column is extracted by investigating the physical relationship of the alphabetic block interpolated in the alphabetic block interpolation section 286. As shown also in the alphabetic block normalization table 381, all alphabetic blocks are expressed as a two-dimensional coordinate which can be determined as a meaning. Therefore, it is possible to extract each line and each train to accuracy by investigating the value of this alphabetic block normalization table 381. For example, it can be supposed that it is one line the alphabetic block which has the value of 3 in the column of Y of the alphabetic block normalization table 381. The processing flow of the line sampling section 287 is shown in drawing 41, and the processing flow of the train extract section 288 is shown in drawing 42.

[0050] In the line sampling section 287, each alphabetic block is first rearranged into ascending order by the Y coordinate in the alphabetic block normalization table 381 (step 4101). Next, the alphabetic block which exists within the limits of height H from the Y coordinate in the range 381 of the lengthwise direction in which an alphabetic block Bi exists the i-th alphabetic block Bi in the alphabetic block which is not registered into ejection (it resets to i= 0 at first) and a line, i.e., the alphabetic block normalization table of an alphabetic block Bi, is registered as one line (step 4102). This processing is repeated until a non-registered alphabetic block is lost.

[0051] Moreover, the train extract section 288 rearranges each alphabetic block into ascending order by the X coordinate in the alphabetic block normalization table 381 similarly (step 4201). Next, the alphabetic block which exists within the limits of width of face W from the X coordinate in the range 381 of the longitudinal direction in which an alphabetic block Bi exists the i-th alphabetic block Bi in the alphabetic block which is not registered into ejection (it resets to i= 0 at first) and a train, i.e., the alphabetic block normalization table of an alphabetic block Bi, is registered as one train (step 4202). This processing is repeated until a non-registered alphabetic block is lost. Furthermore, when a line or a train has division, a line and a train can be extracted by performing processing of repeat line sampling and a train extract to each line and a train until two or more lines or trains are no longer extracted. The result which carried out line sampling to the table of drawing 36 (a) is shown in drawing 43 (a), and the result of a train extract is shown in drawing 43 (b).

[0052] As mentioned above, even if the 4th example is a table which has an abbreviation in the content by extending an alphabetic block so that the edge may be arranged, and normalizing a location and magnitude, it can extract the structure of the row and column in a table to accuracy.

[0053] In addition, an example of the experimental result by the example (the 2nd example) of this invention is shown in drawing 44. Drawing 44 (a) is a subject-copy image, and is the table of the mold with which the vertical ruled line as shown in drawing 2 (e) was omitted. Drawing 44 (b) is the result of processing to the subject-copy image of (a). It turns out that the column of each of tables can be recognized. In addition, character recognition processing is performed about the alphabetic character

part. Although it is a few, the error of character recognition is contained.

[0054]

[Effect of the Invention] Since a table is recognized by burying the gap which extends the alphabetic block which constitutes a table and is between alphabetic blocks according to this invention (claims 1 and 2), the structure can be recognized also about a table with which the ruled line of a table is omitted substantially, and even if the location gap between alphabetic blocks is large, moreover, exact recognition can be performed.

[0055] According to this invention (claim 3), since the information on a ruled line is used for the escape of an alphabetic block, the structure of a table can be extracted more to accuracy. For example, even if it is a table (drawing 20 (a)) with the alphabetic block over two or more division trains like [for identification division / the divided train], the structure of a table can be acquired to accuracy (drawing 20 (c)).

[0056] Since according to this invention (claim 4) the rectangle formed by the ruled line of a table is used in order to grasp the structure of a table, even if it is the table of a complicated configuration, the structure can be extracted to accuracy.

[0057] Since according to this invention (claim 5) an alphabetic block is extended so that the edge may be arranged, and a location and magnitude are normalized, even if it is the table which has an abbreviation in the content, the structure of the row and column in a table can be extracted to accuracy.

[Translation done.]

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] Drawing showing the configuration of the 1st example of this invention

[Drawing 2] Drawing showing the example of the table used in a document

[Drawing 3] Drawing for explaining the pixel lump of an alphabetic character

[Drawing 4] Drawing showing the flow (part) of the alphabetic block extract section

[Drawing 5] Drawing showing the flow (continuation of drawing 4) of the alphabetic block extract section

[Drawing 6] (a) It is drawing to reach and for (b) explain an example as a result of an alphabetic block extract.

[Drawing 7] (a) It is drawing to reach and for (b) explain other examples as a result of an alphabetic block extract.

[Drawing 8] Processing flow drawing of an alphabetic block extension

[Drawing 9] Drawing for explaining alphabetic block distribution

[Drawing 10] (a) It is drawing of ** where it reaches and (b) explains the escape of the longitudinal direction of an alphabetic block.

[Drawing 11] Drawing showing the result of an escape of an alphabetic block

[Drawing 12] Processing flow drawing of the line sampling section

[Drawing 13] Processing flow drawing of the train extract section

[Drawing 14] For (b), (a) is drawing for explaining the result of having extracted the train, as a result of extracting a line.

[Drawing 15] Drawing showing the configuration of the 2nd example of this invention

[Drawing 16] Drawing showing the processing flow (part) of the 1st alphabetic block extension

[Drawing 17] Drawing showing the processing flow (continuation of drawing 16) of the 1st alphabetic block extension

[Drawing 18] (a) It is drawing to reach and for (b) explain an example of the processing result of the 1st alphabetic block extension.

[Drawing 19] (a) It is drawing to reach and for (b) explain other examples of the processing result of the 1st alphabetic block extension.

[Drawing 20] (a), (b), and (c) are drawing for explaining the result of an alphabetic block escape.

[Drawing 21] Drawing showing the configuration of the 3rd example of this invention

[Drawing 22] Drawing showing the processing flow (part) of the rectangle frame extract section

[Drawing 23] Drawing showing the processing flow (continuation of drawing 22) of the rectangle frame extract section

[Drawing 24] (a) It is drawing to reach and for (b) explain a rectangle frame configuration table.

[Drawing 25] Processing flow drawing of the configuration frame extract section

[Drawing 26] (a) It is drawing to reach and for (b) explain the example of a configuration frame.

[Drawing 27] (a) It is drawing in which reaching and showing the result from which (b) extracted the structure of a row and column by the configuration of the 3rd example.

[Drawing 28] Drawing showing the configuration of the 4th example of this invention

[Drawing 29] Flow drawing of the processing which extends an alphabetic block in the vertical direction by the 1st alphabetic block high extension

[Drawing 30] Flow drawing of the processing which extends an alphabetic block in the vertical direction by the 1st alphabetic block width-of-face extension

[Drawing 31] Flow drawing of processing of the 2nd alphabetic block width-of-face extension

[Drawing 32] Flow drawing of processing of the 2nd alphabetic block high extension

[Drawing 33] Drawing showing alphabetic block distribution of the lengthwise direction to the table of drawing 6 (a), and a longitudinal direction

[Drawing 34] It is drawing in which (a) shows the extended result of the width of face of an alphabetic block, and (b) shows the extended result of the height of an alphabetic block.

[Drawing 35] Drawing showing an integrated result by the alphabetic block integrated section

[Drawing 36] For (a), (b) is the example of the table which has an abbreviation in the content, and drawing showing the integrated result by the alphabetic block integrated section to the table of (a).

[Drawing 37] Drawing showing the flow of processing of the alphabetic block normalization section

[Drawing 38] Drawing showing an alphabetic block normalization table

[Drawing 39] Drawing showing the flow of processing of the alphabetic block interpolation section

[Drawing 40] (a) is drawing to show the alphabetic block normalization table which can be done as a result of processing of the alphabetic block interpolation section, and for (b) explain the result of processing of the alphabetic block interpolation section.

[Drawing 41] Drawing showing the flow of processing of the line sampling section

[Drawing 42] Drawing showing the flow of processing of the train extract section

[Drawing 43] For (b), (a) is drawing for explaining the result of a train extract as a result of line sampling.

[Drawing 44] It is drawing showing the result to which (a) processed (b) to the subject-copy image to the subject-copy image of a table.

[Description of Notations]

11,151,211,281 -- An alphabetic character / ruled line separation section, 12,152,212,282 -- Alphabetic block extract section, a 13,153,215 -- alphabetic block extension, 1531 and 2151, and -- the 1st alphabetic block extension -- 1532 2152 -- 14,154,216 The 2nd alphabetic block extension, 287 -- Line sampling section, 15,155,217 288 -- The train extract section, 214 -- Configuration frame extract section, 2831 -- The 1st alphabetic block width-of-face extension, 2832 -- The 1st alphabetic block high extension, 2841 [-- The alphabetic block normalization section, 2851 / -- The alphabetic block width-of-face normalization section, 2852 / -- The alphabetic block high normalization section, 286 / -- Alphabetic block interpolation section] -- The 2nd alphabetic block width-of-face extension, 2842 -- The 2nd alphabetic block high extension, 2843 -- The alphabetic block integrated section, 285

[Translation done.]

* NOTICES *

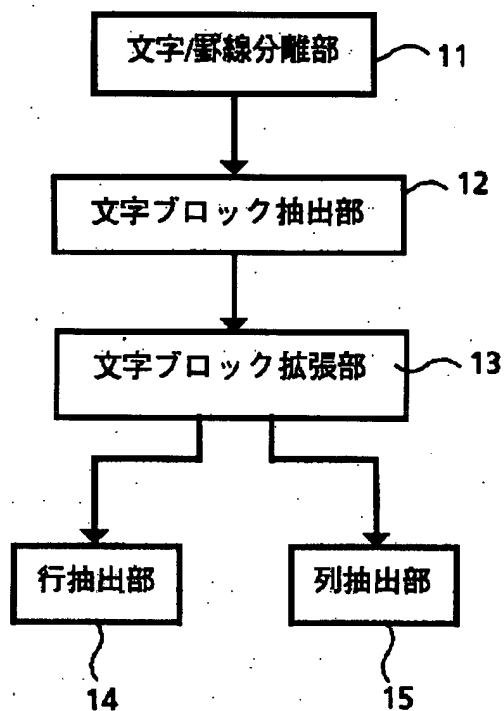
JPO and NCIPi are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DRAWINGS

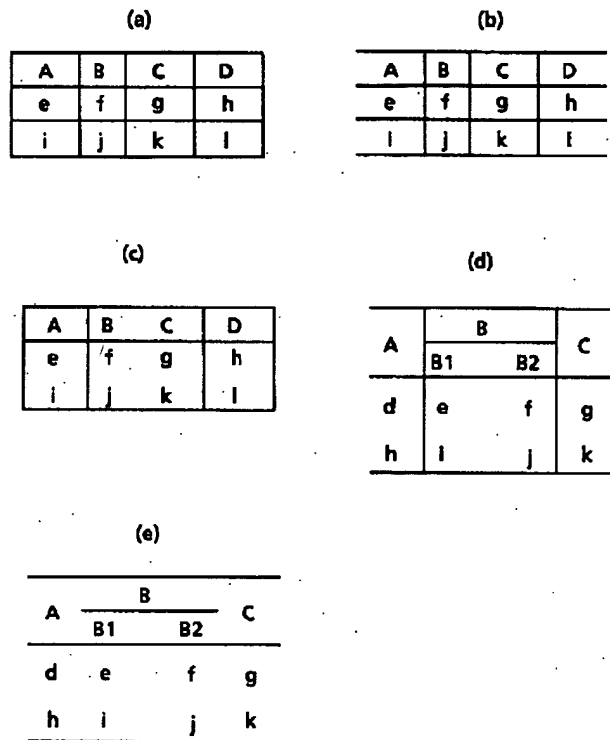
[Drawing 1]

図1



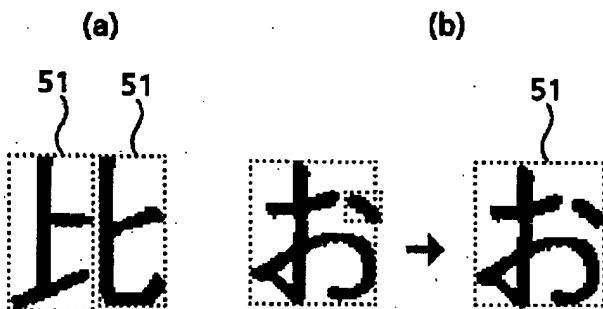
[Drawing 2]

図2



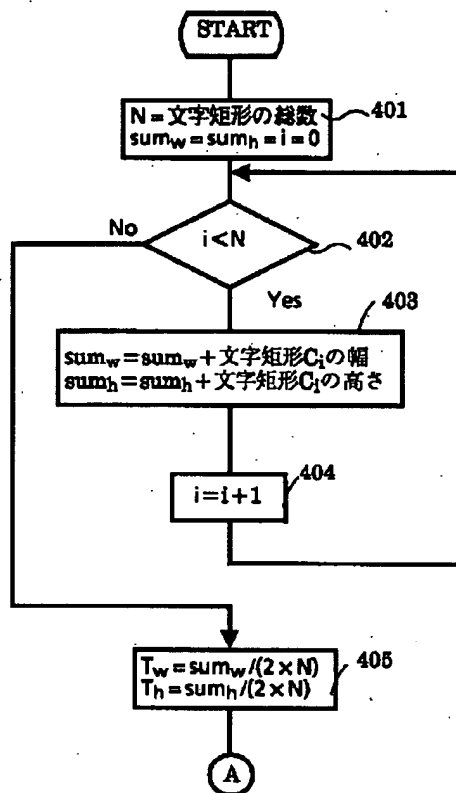
[Drawing 3]

図3



[Drawing 4]

図4



[Drawing 6]

図6

(a)

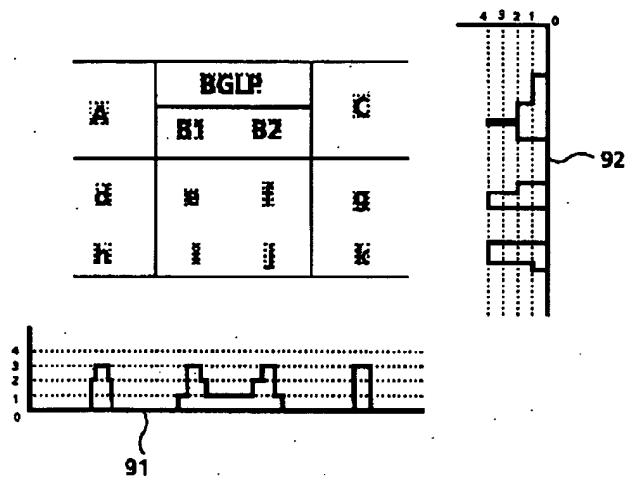
A	BGLP		C
	B1	B2	
d	e	f	g
h	i	j	k

(b)

61	BGLP		C
	B1	B2	
62	63	64	65
66	67	68	69

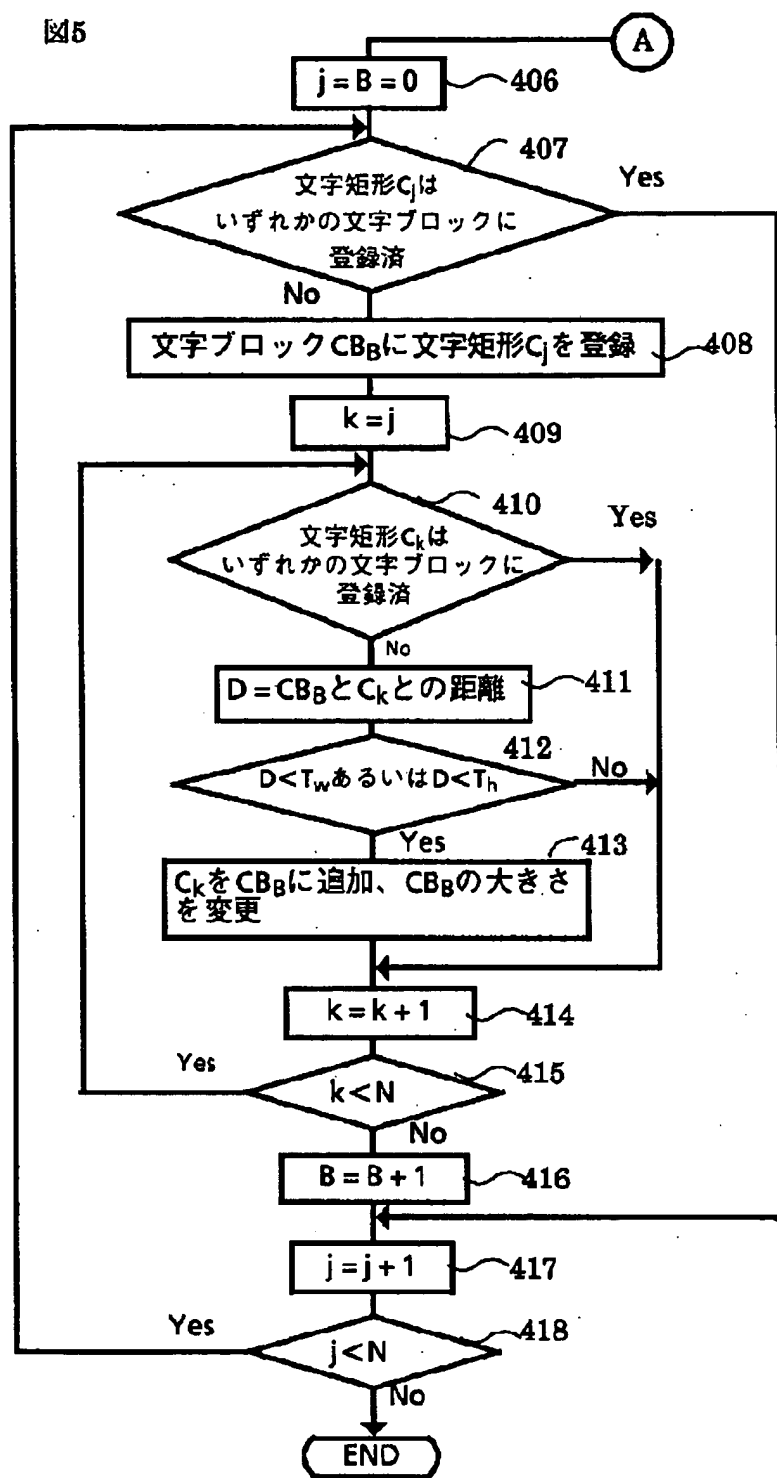
[Drawing 9]

図9



[Drawing 5]

図5

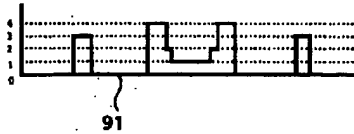


[Drawing 10]

FIG. 10

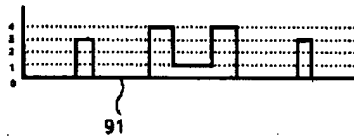
(a)

A	BGLP		C
	B1	B2	
d	e	f	g
h	i	j	k



(b)

A	BGLP		C
	B1	B2	
d	e	f	g
h	i	j	k



[Drawing 7]

図7

(a)

昨日	今日	明日
富士	赤坂	六本木
武蔵小金井	渋谷	新宿

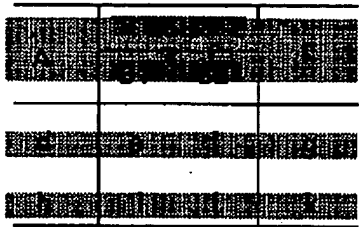
(b)

昨日	今日	明日 61
富士	赤坂	六本木
武蔵小金井	渋谷	新宿

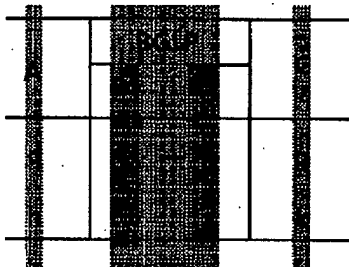
[Drawing 14]

図14

(a)

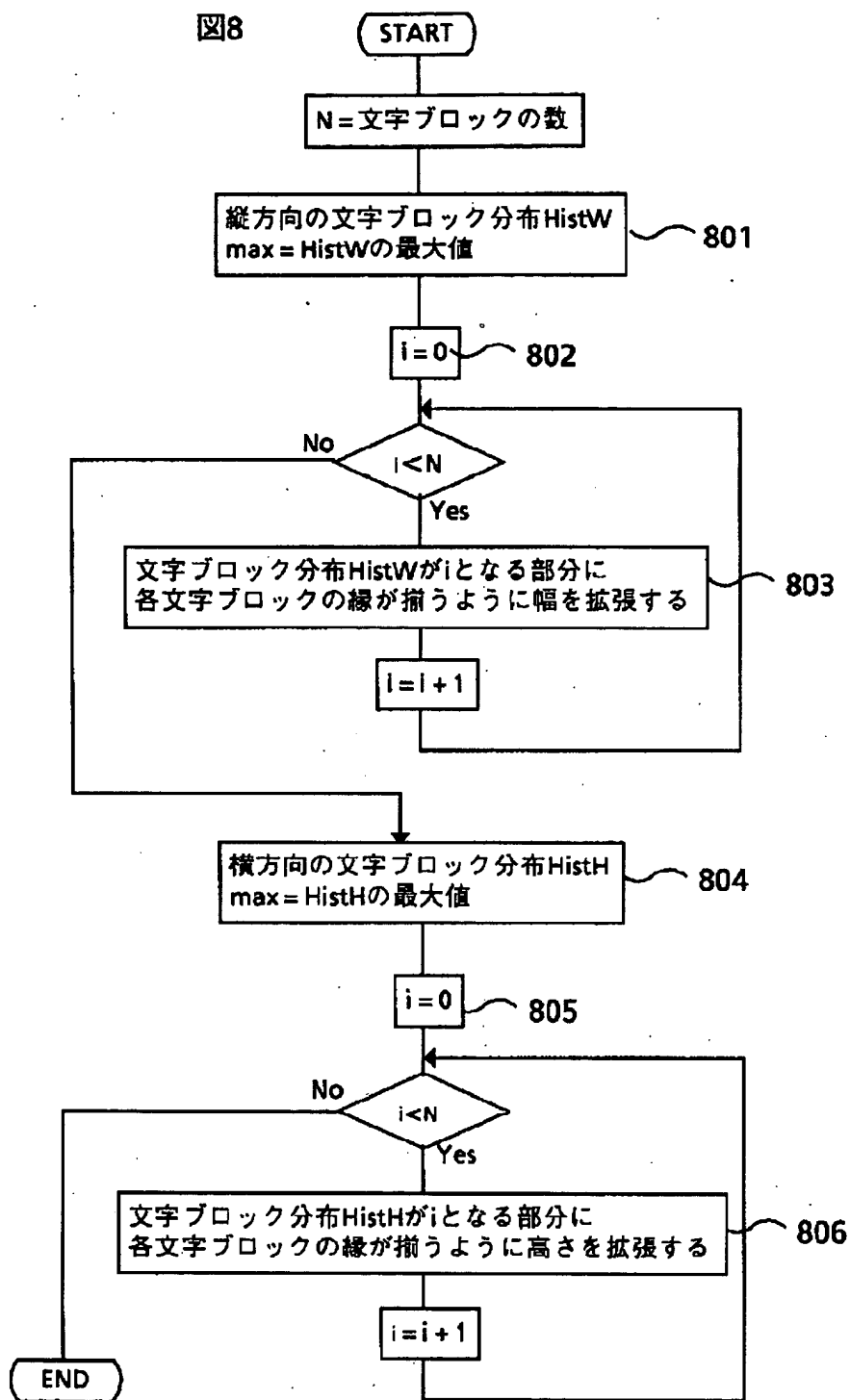


(b)



[Drawing 8]

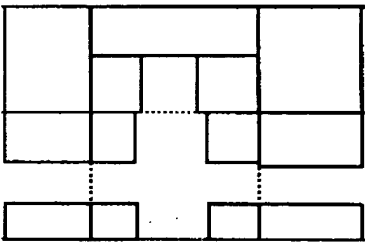
図8



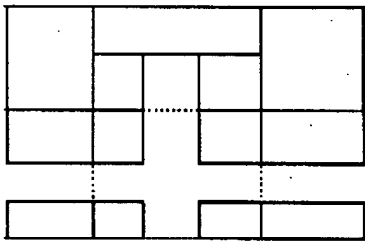
[Drawing 18]

図18

(a)

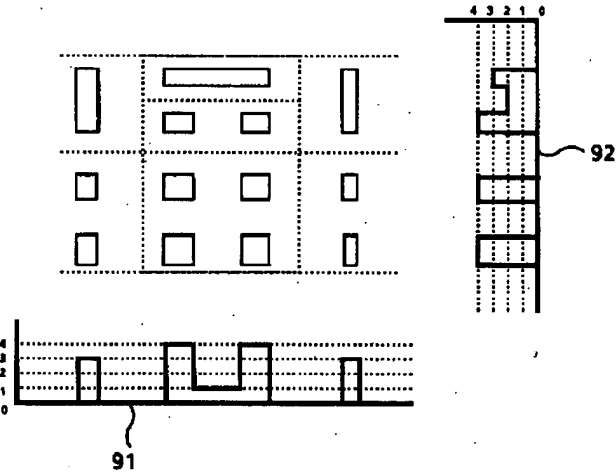


(b)



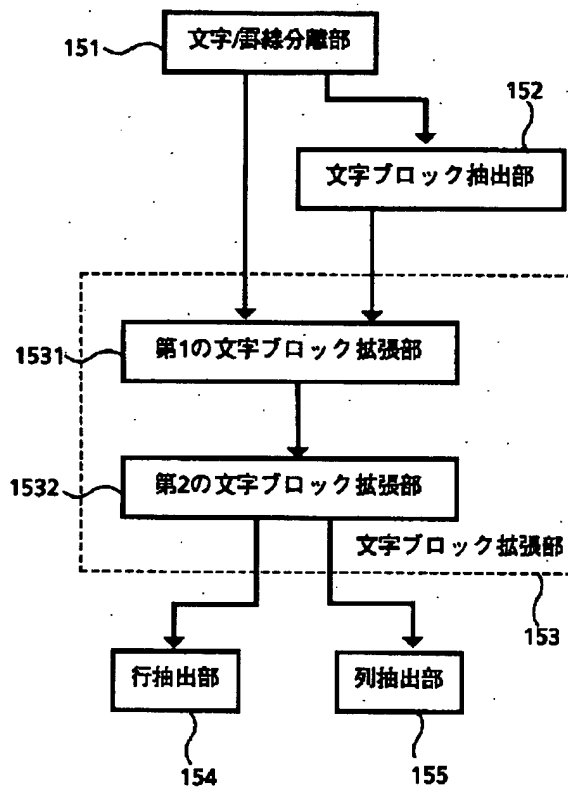
[Drawing 11]

図11



[Drawing 15]

図15



[Drawing 19]

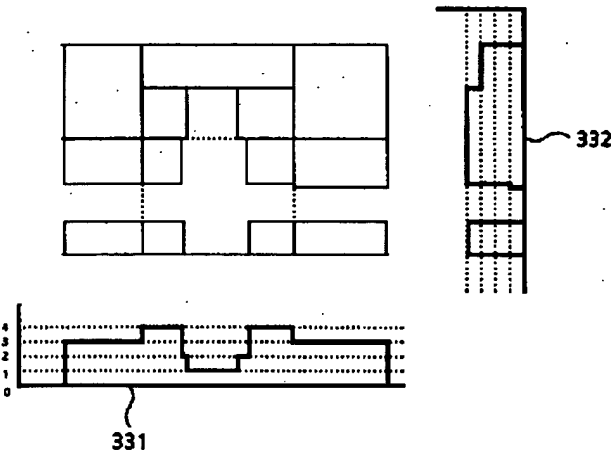
図19

(a)

A	B	C	D
EF	GH	KL	MN
OP	QR	ST	UV

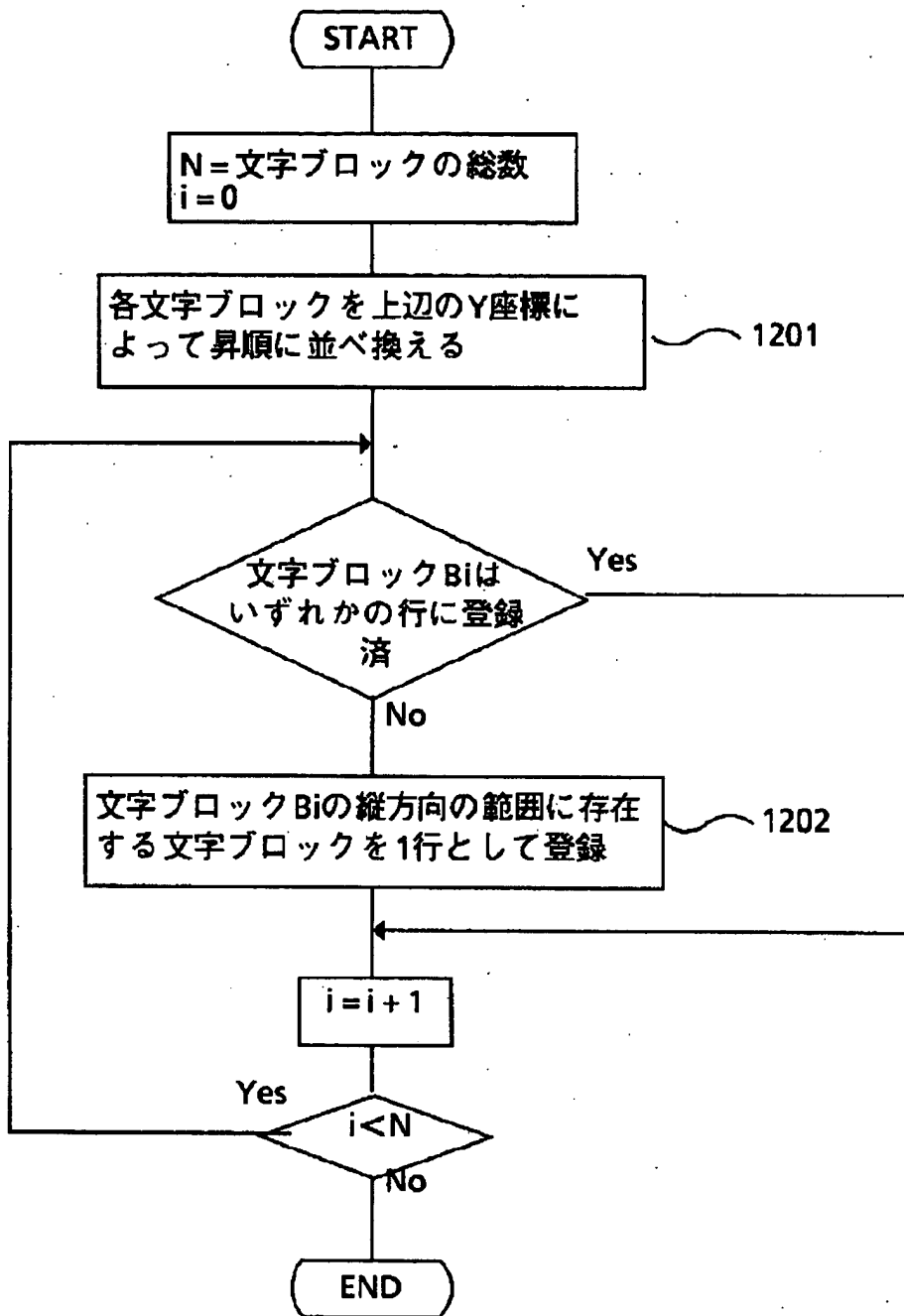
(b)

[Drawing 33] 図33



[Drawing 12]

図12



[Drawing 26]

図26

(a)

A	BGLP		C
	B1	B2	
D	E	F	G
H	I	J	K

(b)

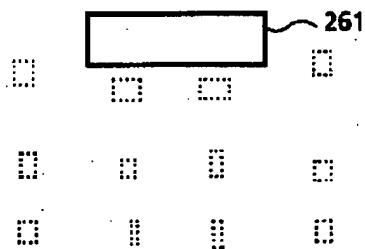
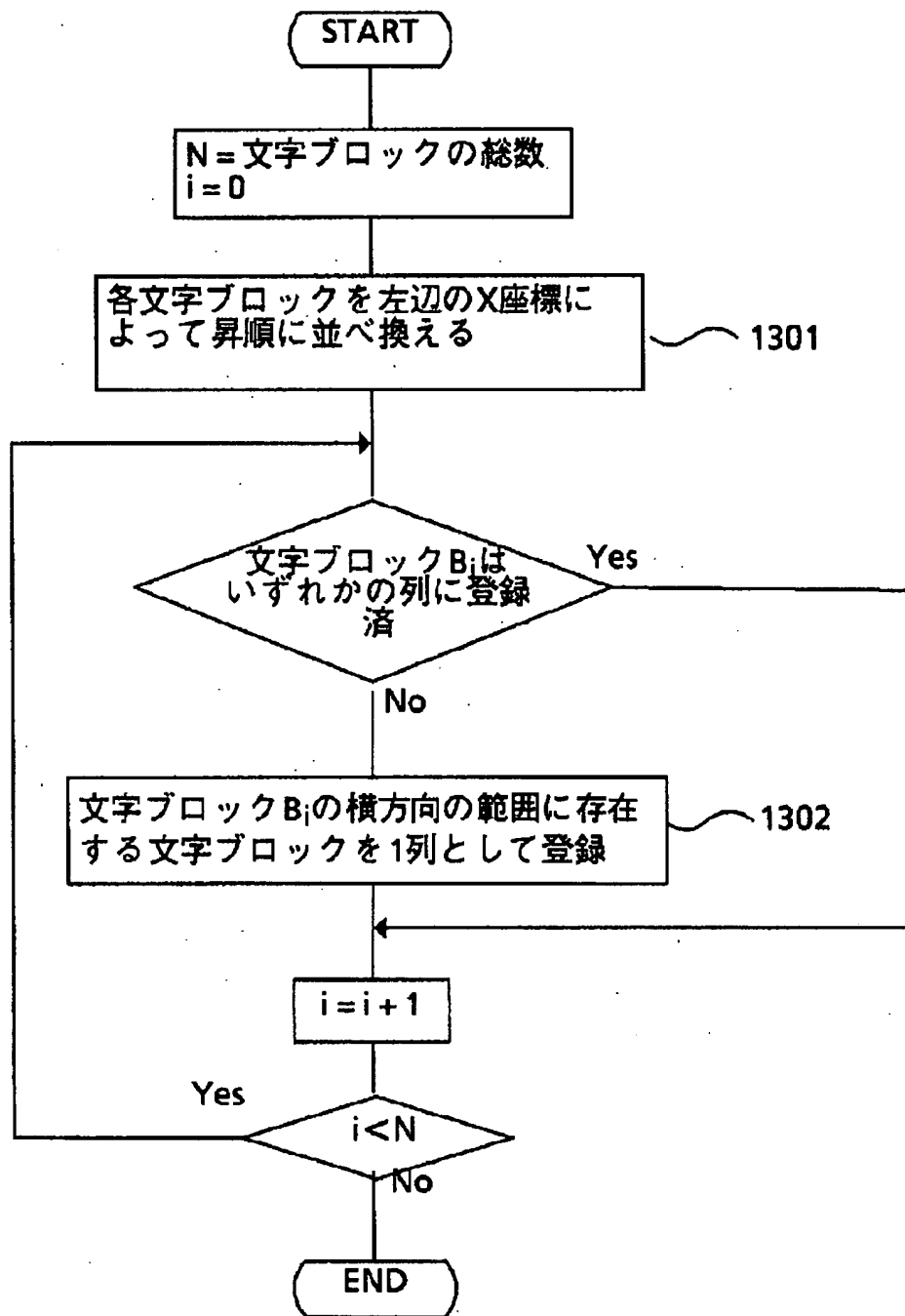
[Drawing 13]

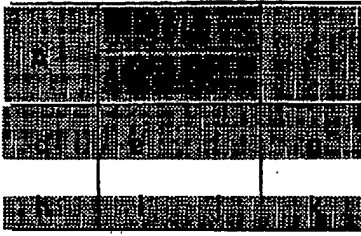
図13



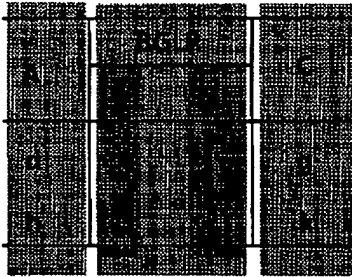
[Drawing 27]

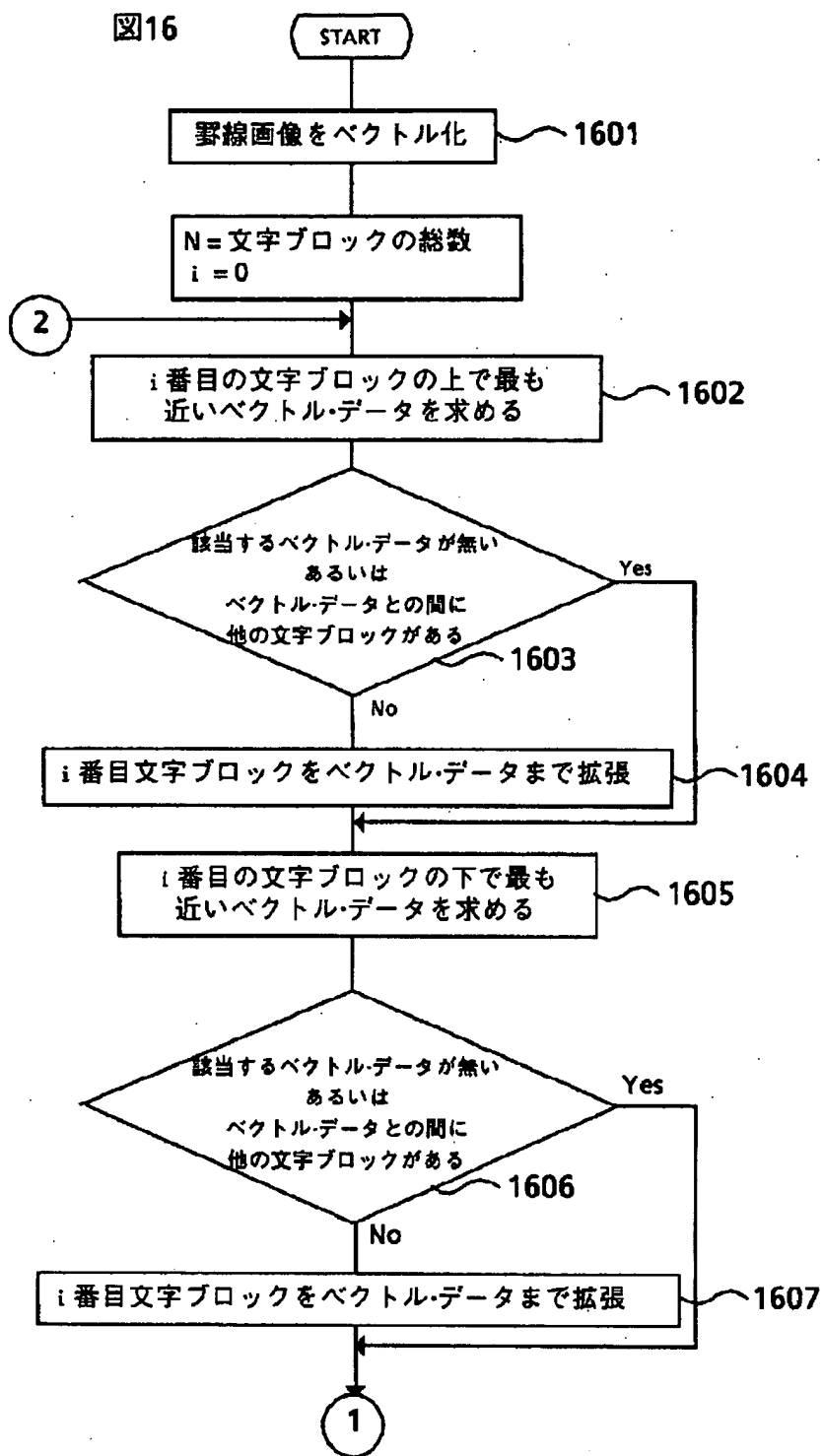
図27

(a)



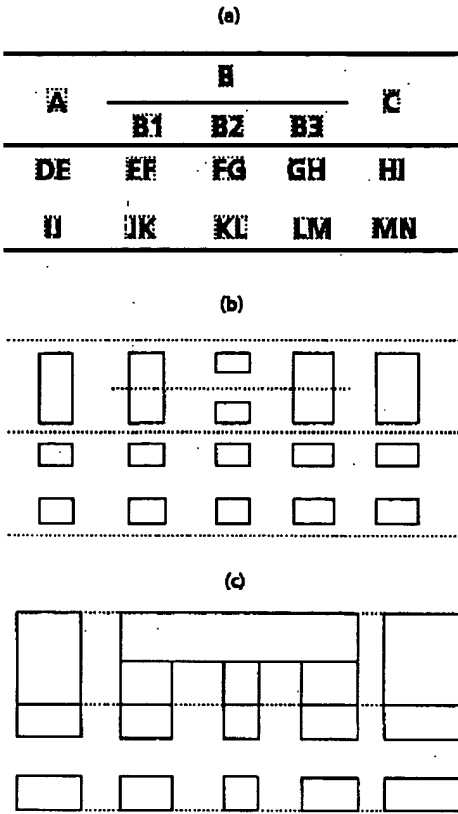
(b)

[Drawing 16]



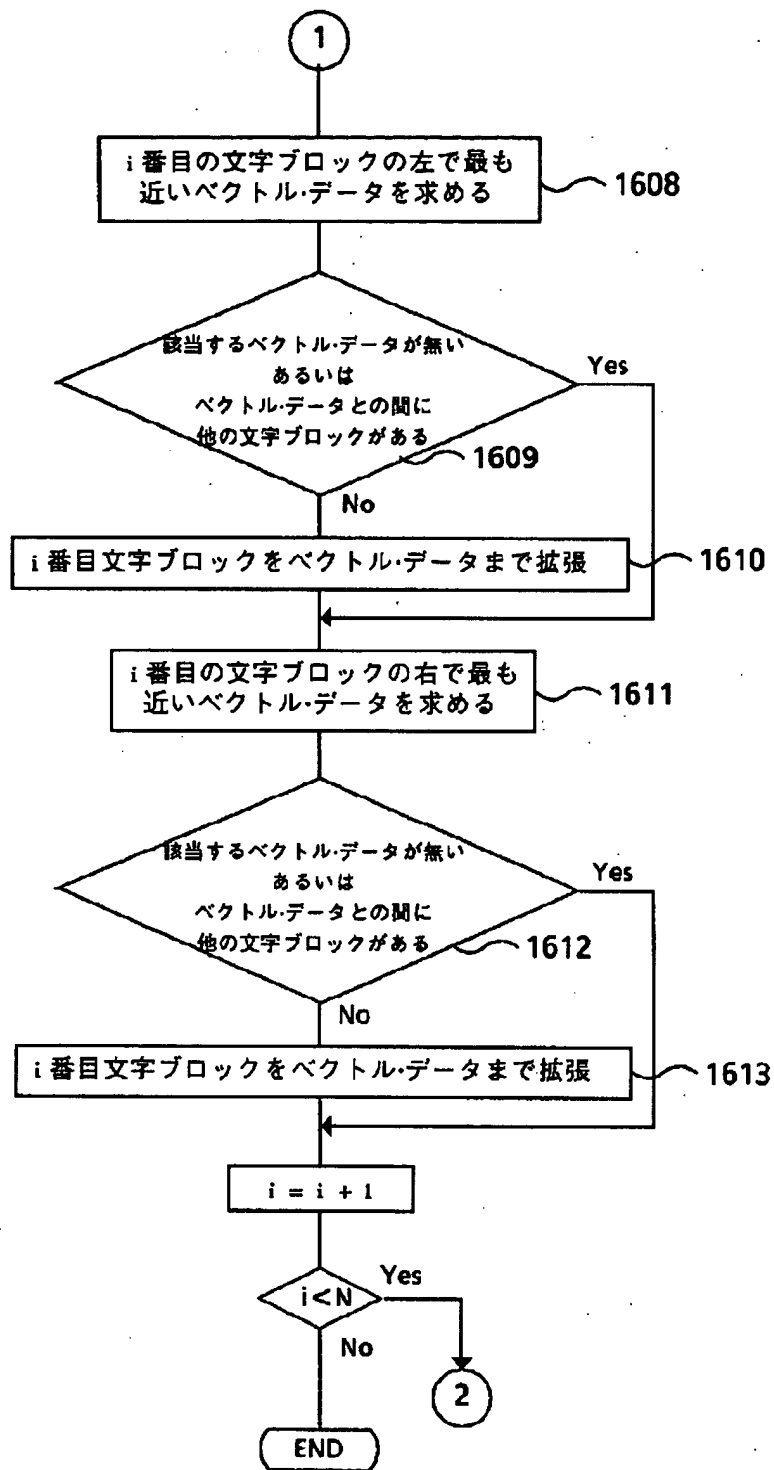
[Drawing 20]

FIG. 20



[Drawing 17]

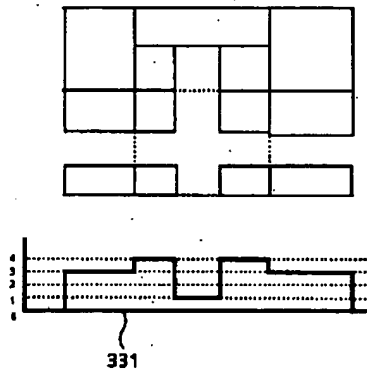
図17



[Drawing 34]

Fig. 34

(a)



(b)

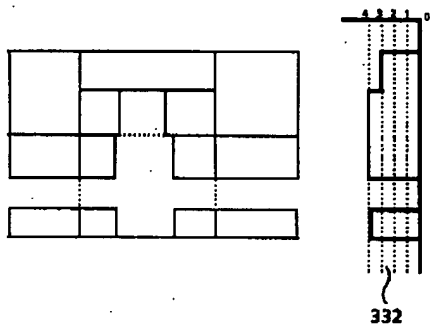
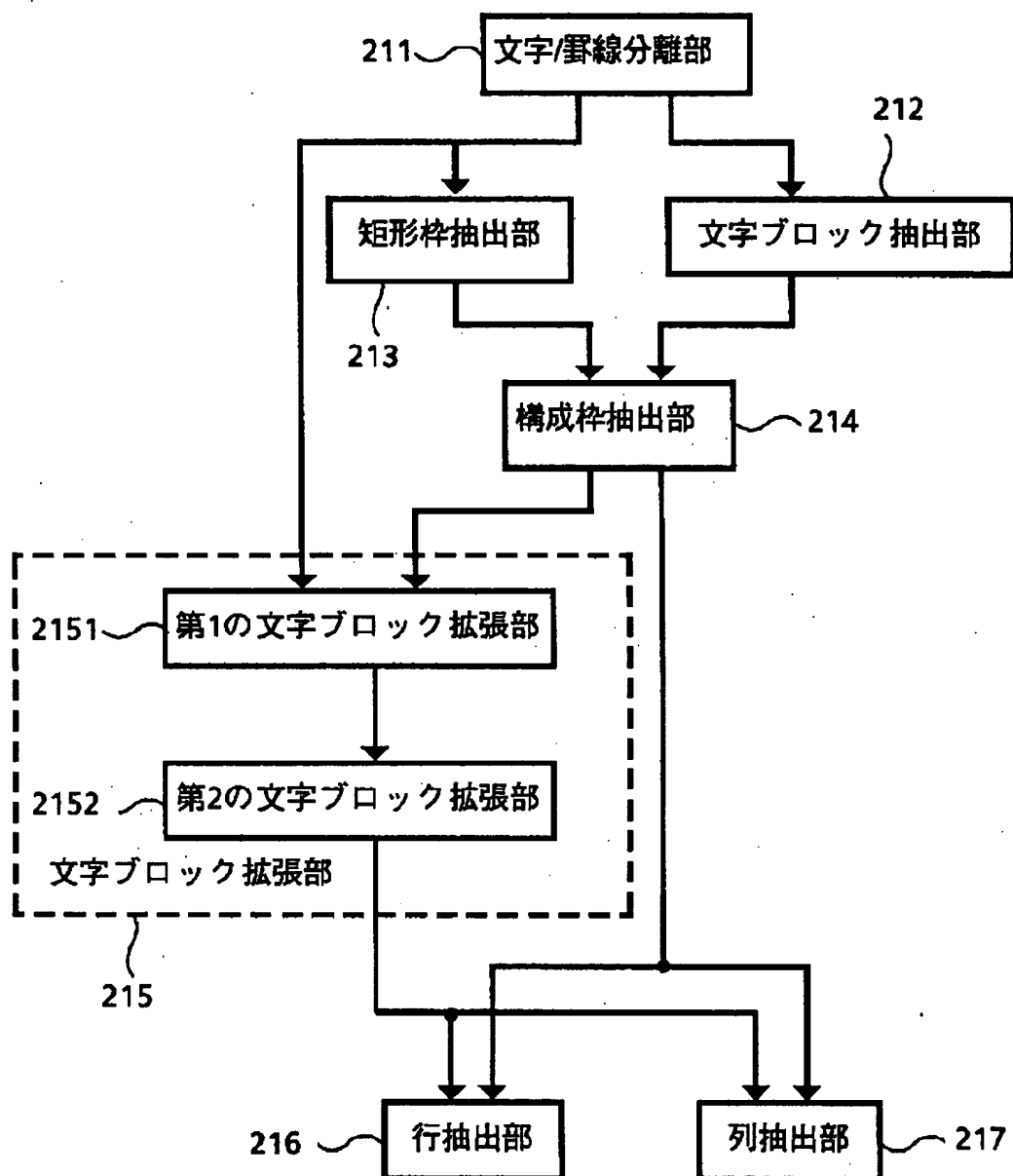
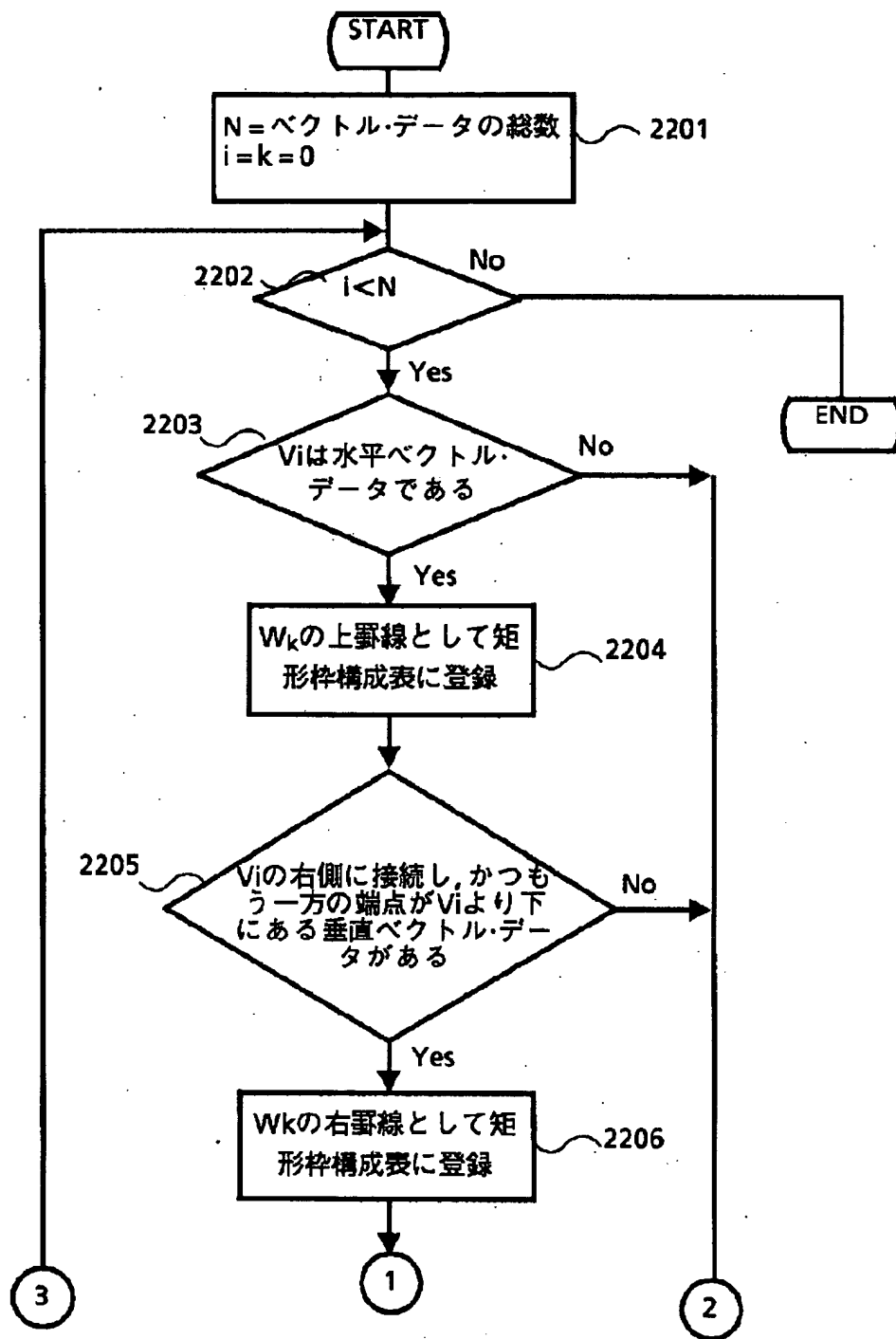
[Drawing 21]

図21



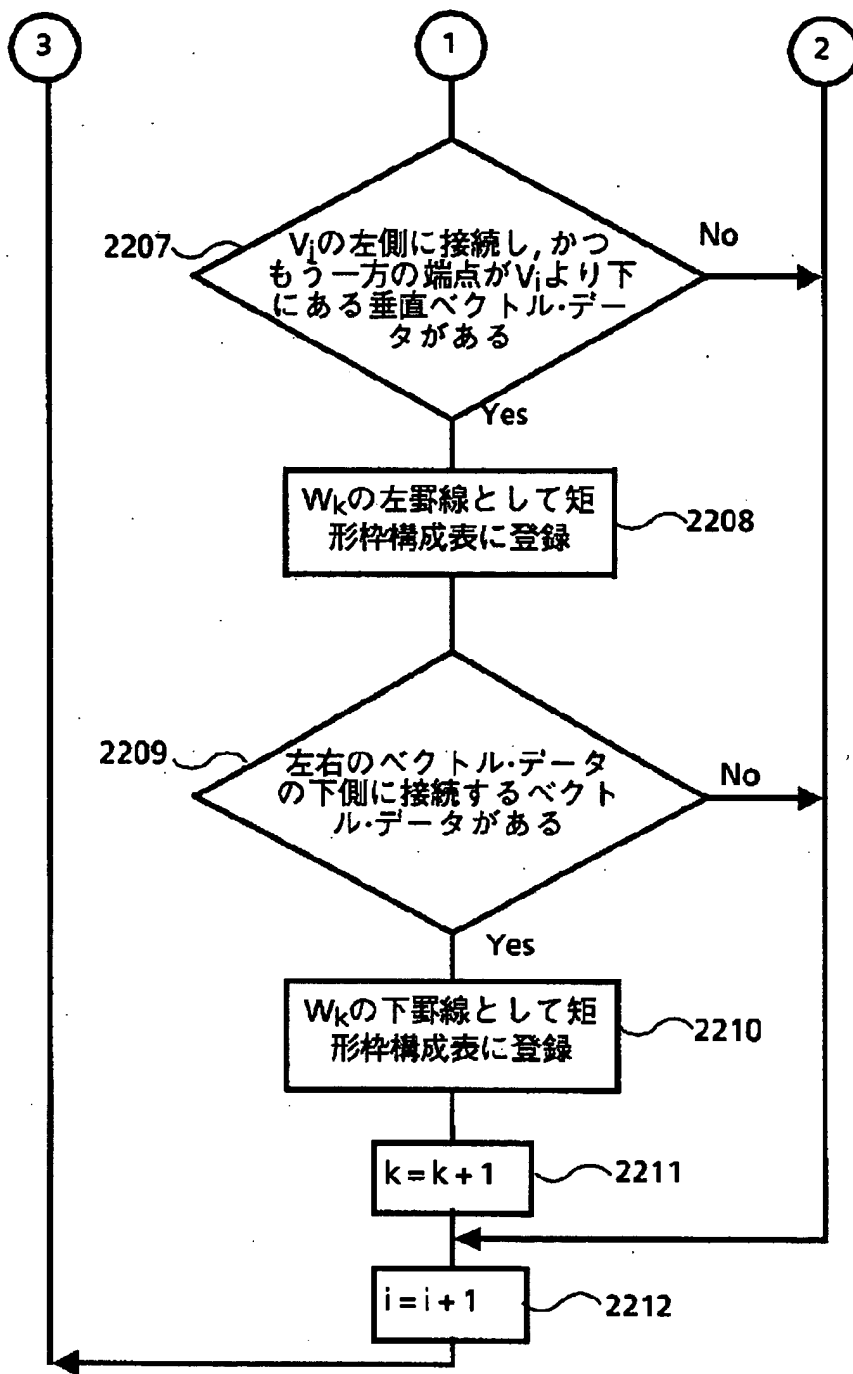
[Drawing 22]

図 22



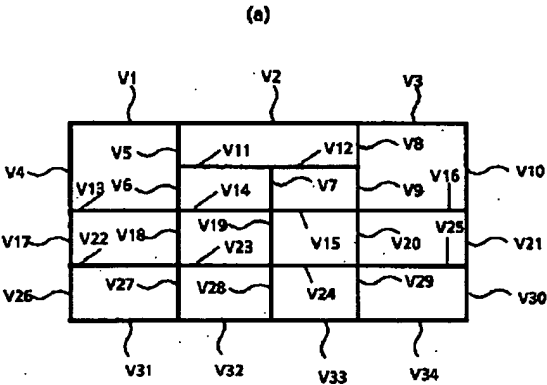
[Drawing 23]

図 23



[Drawing 24]

図24



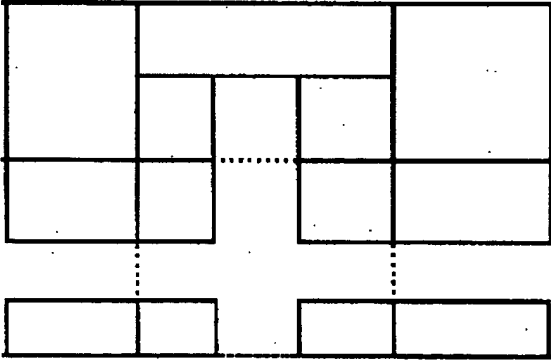
(b)

241

枠No	上野線	右野線	左野線	下野線
W ₀	V1	V5,6	V4	V13
W ₁	V2	V8	V5	V11,12
W ₂	V11	V7	V6	V14
W ₁₂	V25	V30	V29	V34

[Drawing 35]

図35



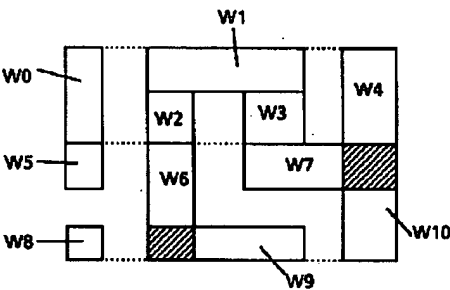
[Drawing 36]

図36

(a)

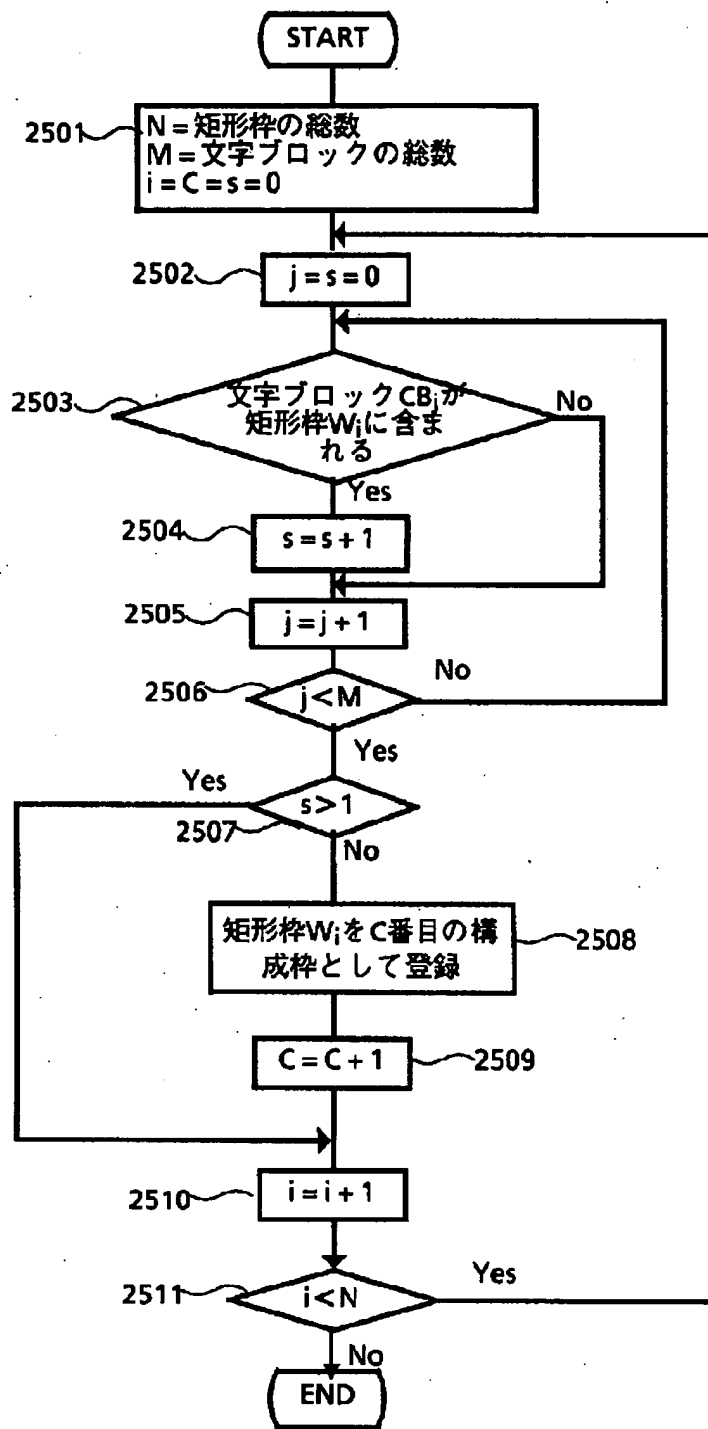
BGLP			
A	B1	B2	C
d	e	f	
h		j	k

(b)



[Drawing 25]

図25



[Drawing 40]

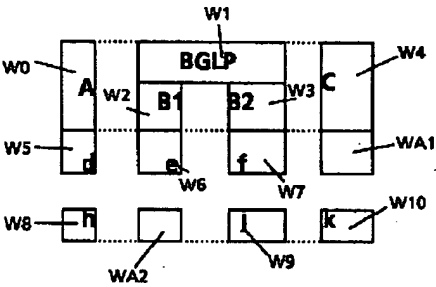
図40

(a)

381

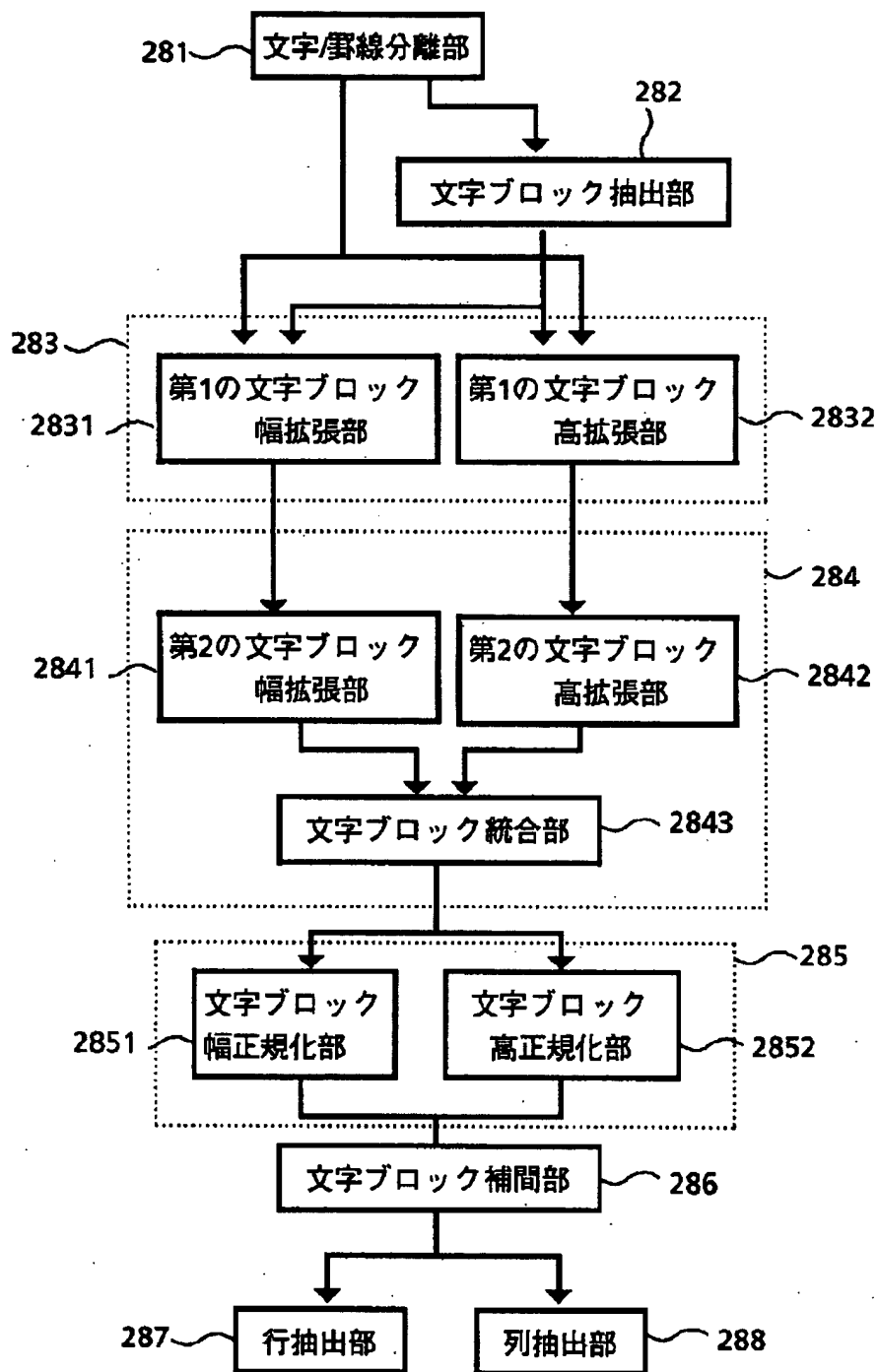
枠No	X	W	Y	H
W ₀	1	1	1	2
W ₁	2	2	1	1
W ₂	2	1	2	1
W ₁₀	4	1	4	1
WA ₁	4	1	3	1
WA ₂	2	1	4	1

(b)



[Drawing 28]

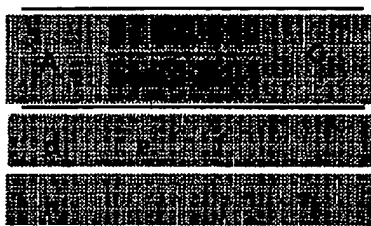
図28



[Drawing 43]

図43

(a)



(b)

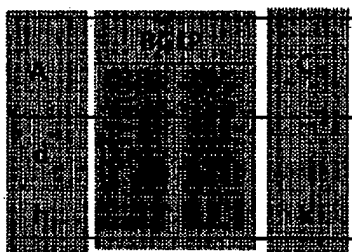
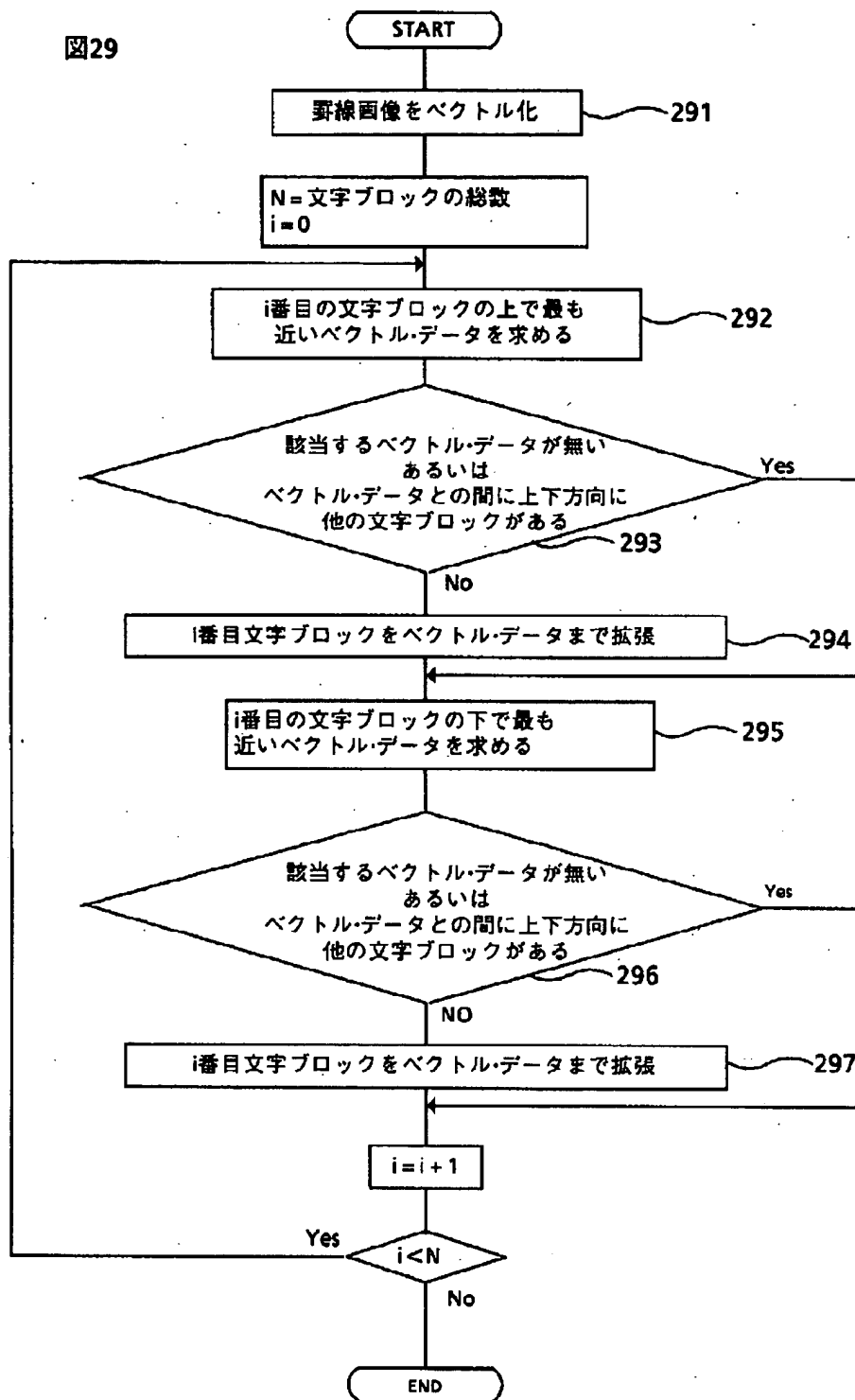
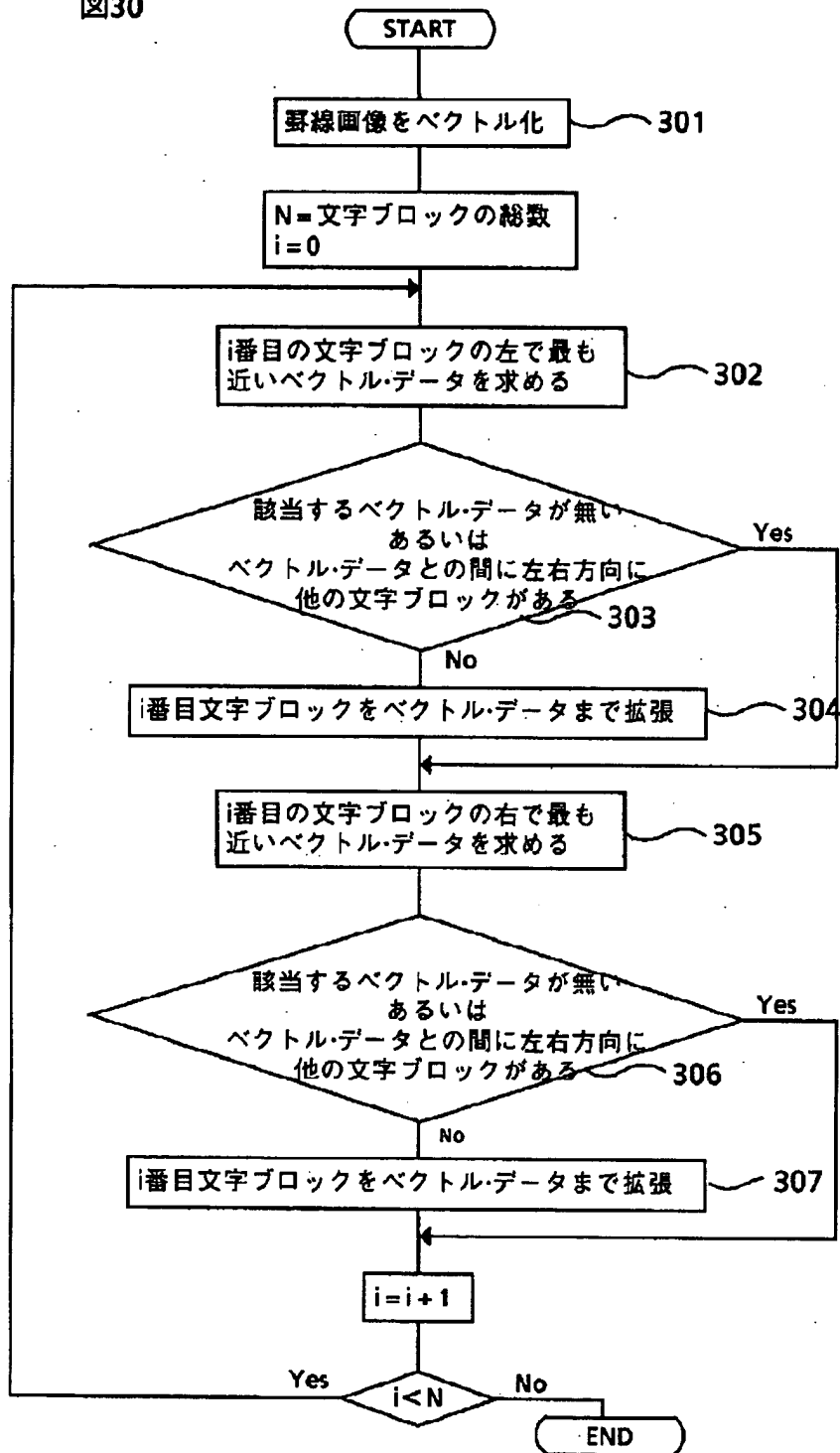
[Drawing 29]

図29



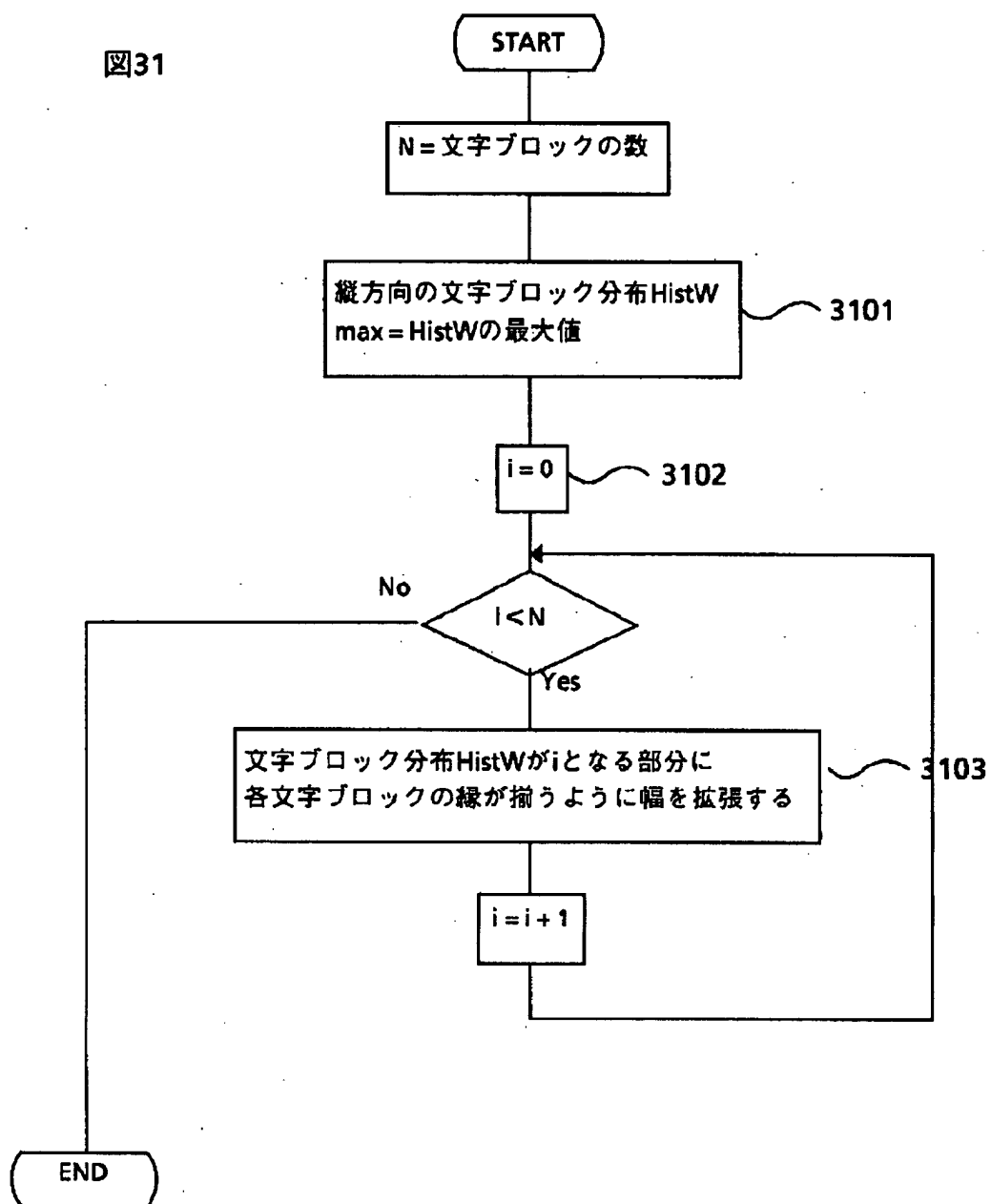
[Drawing 30]

図30



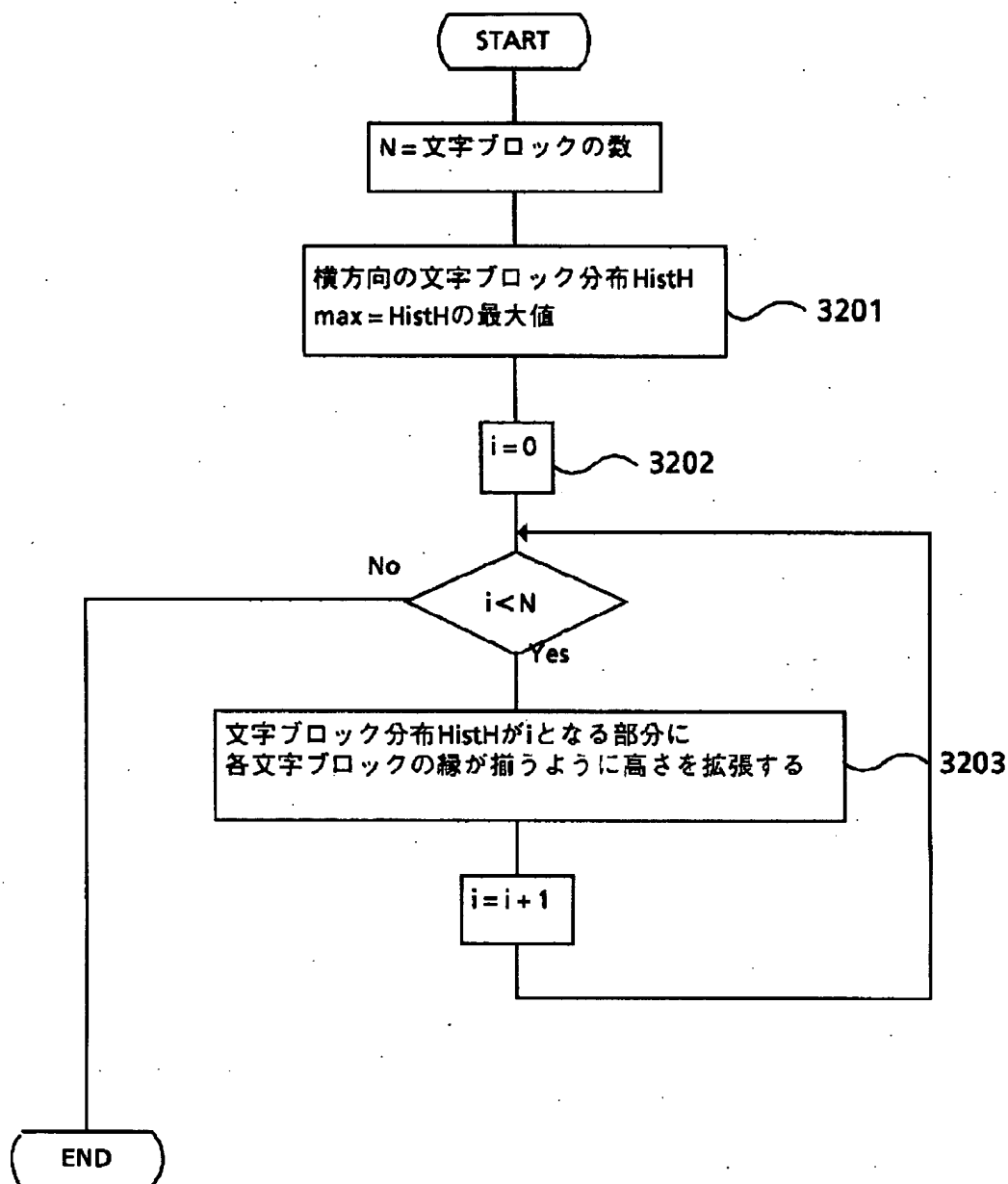
[Drawing 31]

図31



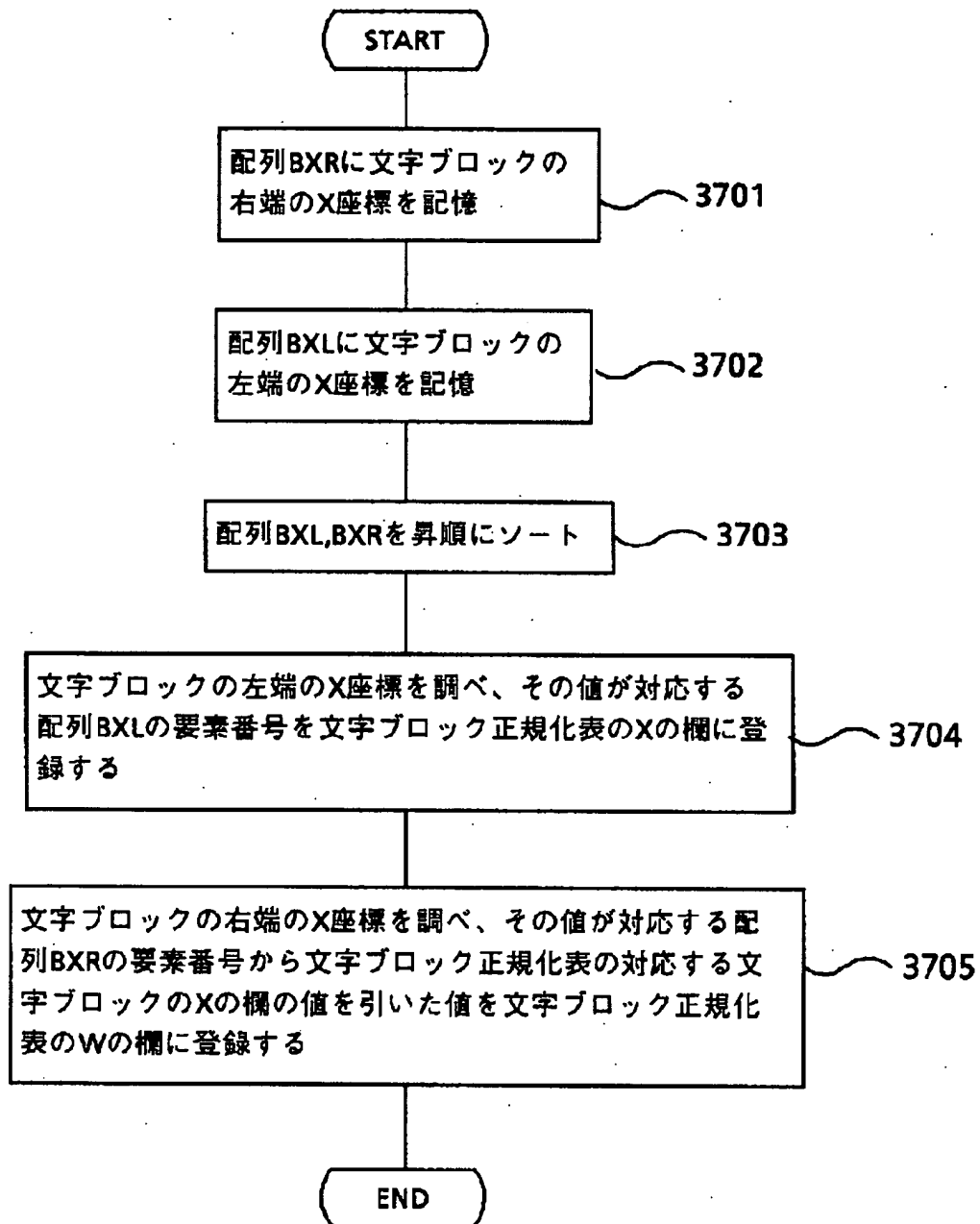
[Drawing 32]

図32



[Drawing 37]

図37



[Drawing 38]

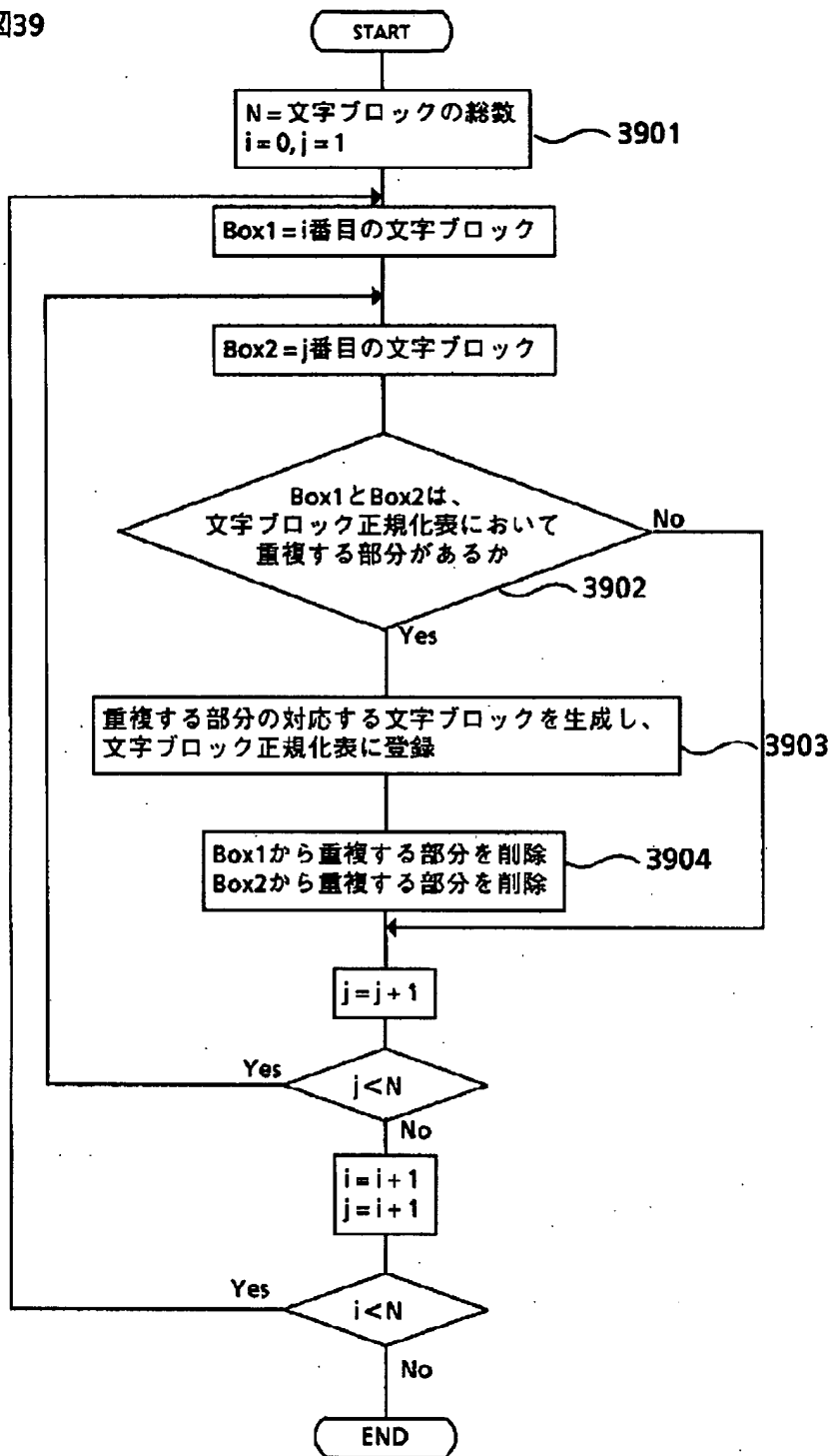
図38

381

枠No	X	W	Y	H
W ₀	1	1	1	2
W ₁	2	2	1	1
W ₂	2	1	2	1
W ₁₀	4	1	3	2

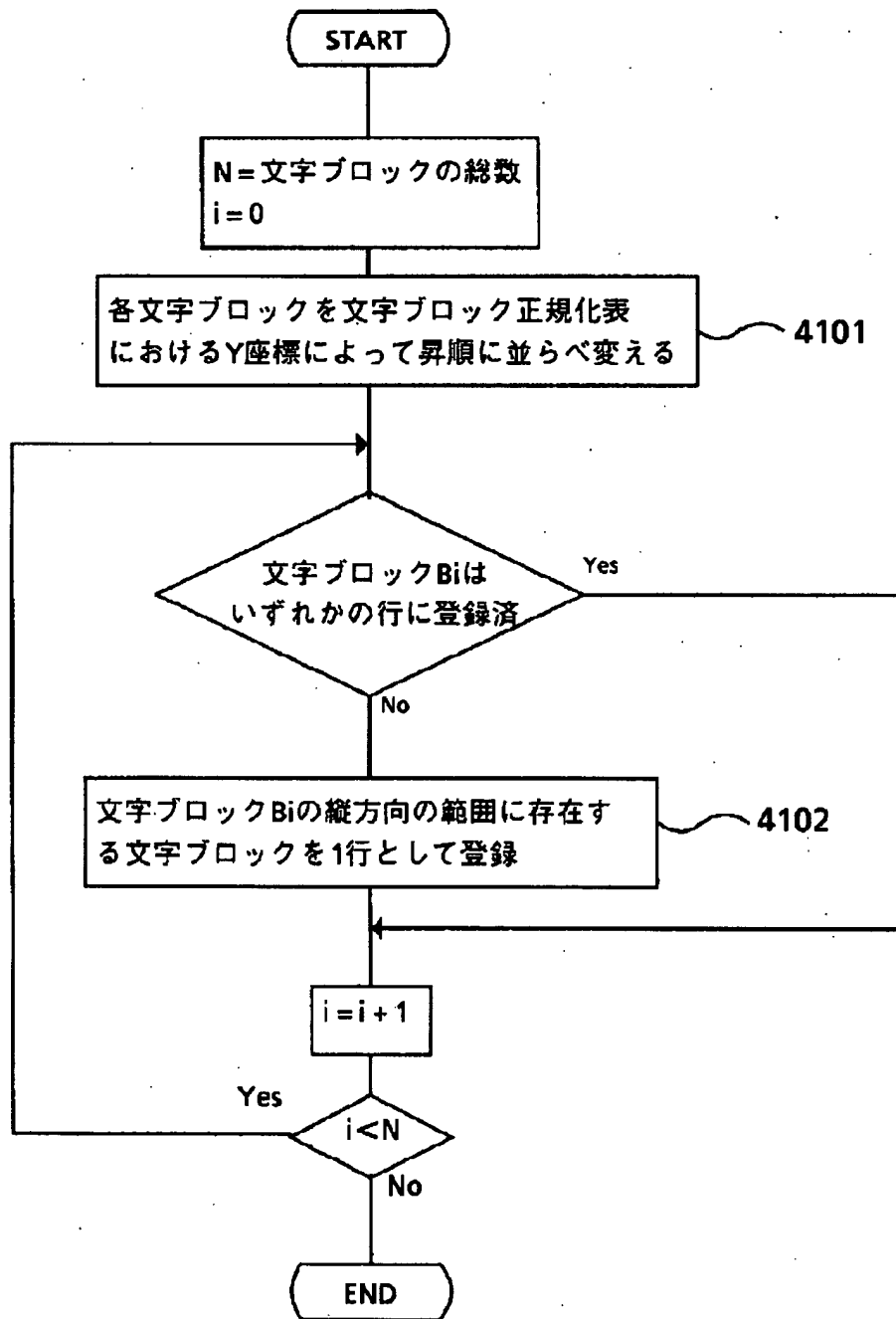
[Drawing 39]

図39



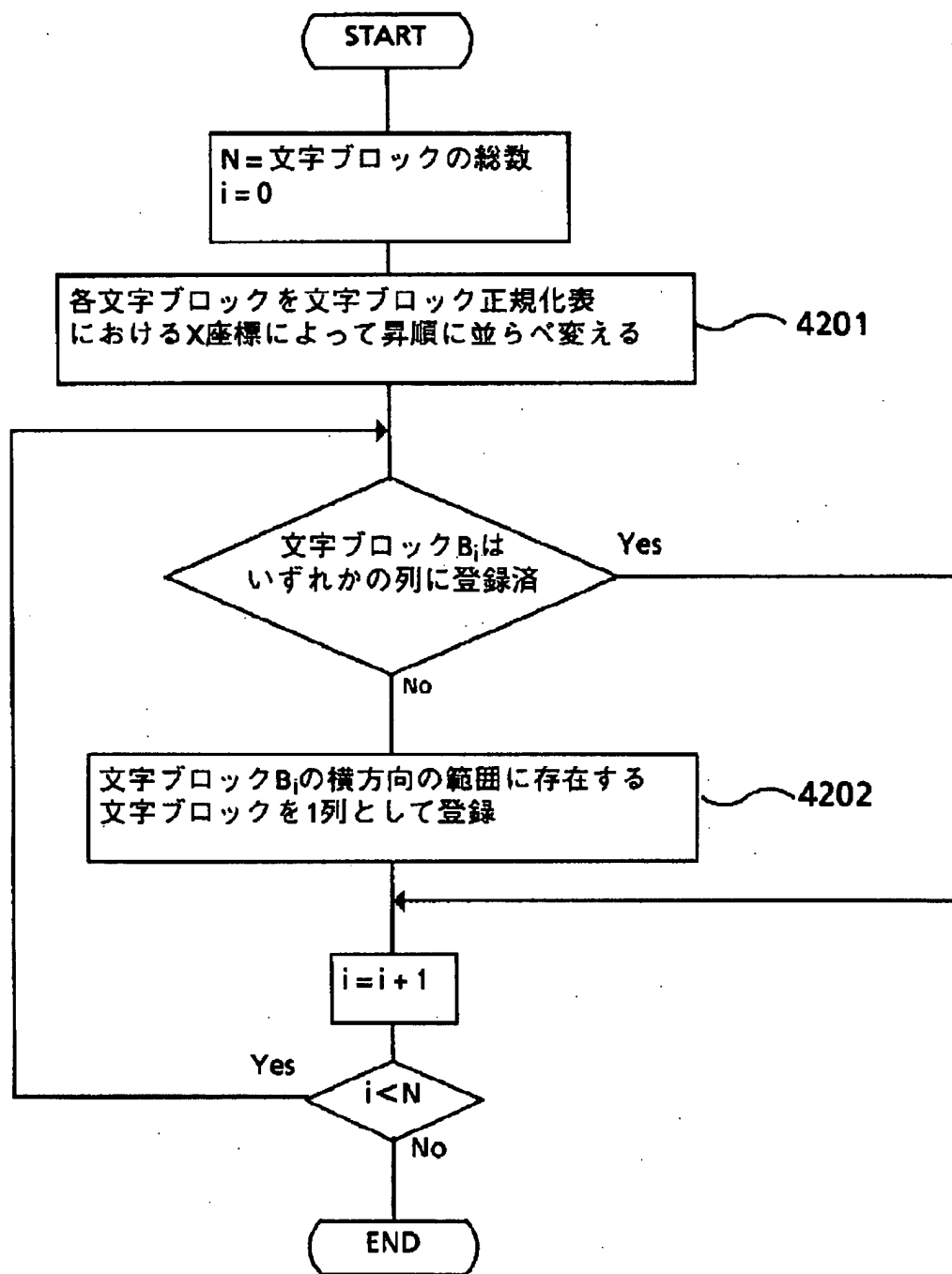
[Drawing 41]

図41



[Drawing 42]

図42



[Drawing 44]

図44

(a)

Condition	Diffuses illutin				Spot			conductance
	IPL	Sumi	Vitreol	Simu	IPL	PNR	wave	
Control	246	211	45	15%	484	451	407	92
Control	213	86%	34	17	346	378	381	22%

(b)

Condition	diffuses illutin					Spot			conductance
	PL	Sum	Vitreol	Simu	PL	PNR	wave		
Control	246	211	45	15%	484	451	407	92	
Control	213	86%	34	17	346	378	381	22%	

[Translation done.]